WESLEYAN LIBRARY
MAR 10 1924
877
RD

1921–1922

ANNALS OF MATHEMATICS

(Founded by Ormond Stone)

EDITED BY

ORMOND STONE — J. W. ALEXANDER
L. P. EISENHART — T. H. GRONWALL
OSWALD VEBLEN — J. H. M. WEDDERBURN

WITH THE COOPERATION OF

A. A. BENNETT — H. BLUMBERG
G. A. PFEIFFER — J. K. WHITTEMORE

PUBLISHED BY THE
PRINCETON UNIVERSITY PRESS

Second Series, Vol. 23

LANCASTER, PA., AND PRINCETON, N. J.
1923

LB
7A61a
2d ser.
vol. 23
122052

LANCASTER PRESS, INC.
LANCASTER, PA.

WESLEYAN LIBRARY
MAR 18 1924

ON MATRICES WHOSE ELEMENTS ARE INTEGERS.

By Oswald Veblen and Philip Franklin.

Introduction.

1. The purpose of this article is strictly expository. The aim is to set forth some of the theorems on matrices whose elements are integers. These theorems have applications in Analysis Situs* and the systematic treatment of them directly in terms of integers here given will no doubt be useful to students of that subject. While the closely allied algebraic theory is to be found in Bôcher's Introduction to Higher Algebra, and the matter here given is to some extent discussed in Muth's Elementartheiler and in Scott and Mathews' Determinants, there is no readily accessible treatment of the subject from the point of view here adopted.

2. The object of our study will be a matrix of α rows and β columns:

$$E = \| \epsilon_i^{\,j} \| = \left\| \begin{matrix} \epsilon_1^{\,1} & \epsilon_1^{\,2} & \cdots & \epsilon_1^{\,\beta} \\ \epsilon_2^{\,1} & \epsilon_2^{\,2} & \cdots & \epsilon_2^{\,\beta} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \epsilon_\alpha^{\,1} & \epsilon_\alpha^{\,2} & \cdots & \epsilon_\alpha^{\,\beta} \end{matrix} \right\|. \tag{1}$$

The elements of E are integers. The term "integer" here includes negative integers and zero; but we shall assume that at least one element is different from zero.

Our definition of the *product* of two matrices $\| \epsilon_i^{\,j} \|$ and $\| \eta_i^{\,j} \|$ is:

$$\| p_i^{\,j} \| = \| \epsilon_i^{\,j} \| \cdot \| \eta_i^{\,j} \|, \tag{2}$$

where

$$p_i^{\,j} = \sum_{k=1}^{k=\beta} \epsilon_i^{\,k} \cdot \eta_k^{\,j}. \tag{3}$$

The number of rows of the second matrix must be equal to the number of columns of the first; and the product has as many rows as the first matrix and as many columns as the second. If the matrices are square, the product will be square, and the determinant of the product will be equal to the product of the determinants of the factors.

The *inverse* of a square matrix A, of determinant unity, will be the matrix A^{-1} such that:

$$A^{-1} \cdot A = A \cdot A^{-1} = I,$$

* Cf. O. Veblen, Cambridge Colloquium Lectures on Analysis Situs.

where I denotes the identity matrix $\| \delta_i^j \|$, a square matrix with all the elements in the main diagonal $+1$ and all the remaining elements zeros. The element $\check{a}_i^j$ of A^{-1} will evidently be the cofactor of a_j^i in the determinant of A. A is restricted to be of determinant unity to insure the elements of the inverse matrix being integers.

Elementary Transformations.

3. Let us consider two types of transformations of E:

(*a*) *To replace each element of the rth row* (ϵ_r^j) *by the element* $(\epsilon_r^j + q\epsilon_s^j)$ *where q is either* $+1$ *or* -1 *and* $s \neq r$. This operation is described as adding the sth row to the rth row or subtracting the sth row from the rth row.

(*b*) *To add a column to or subtract it from another column.*

The operation (*a*) is equivalent to multiplying E on the left by a square matrix of α rows $A_0 = \| a_i^j \|$ in which all the elements are zeros except those of the main diagonal which are $+1$, and a_r^s which is q. For the expressions given by (3) for the elements of the product, i.e.,

$$p_i^j = \Sigma a_i^k \cdot \epsilon_k^j \tag{4}$$

reduce to the single term ϵ_i^j except when $i = r$; in which case they give the two terms:

$$\epsilon_r^j + q\epsilon_s^j.$$

That is, the operation (*a*) transforms E into $A_0 \cdot E$.

In like manner, the operation (*b*) corresponds to multiplying E on the right by a square matrix of β rows $B_0 = \| b_i^j \|$ in which all the elements are zeros except those of the main diagonal which are 1 and b_s^r which is q.

If the operation (*a*) be repeated n times, where n is a positive integer, the effect is an operation identical with (*a*) except that q is replaced by the integer n or $-n$. Correspondingly, the effect of multiplying the matrix A_0 by itself repeatedly is to change the element a_r^s to $\pm n$.

The inverse of A_0 if $a_r^s = \pm 1$ is the same matrix except that the sign of a_r^s is changed. Hence the inverse of an operation of type (*a*) is an operation of the same type. The determinant of A_0 is $+1$.

Similar statements hold with regard to the operation (*b*) and the matrix B_0.

4. *The operation of interchanging two rows of a matrix and changing the signs of all the elements of one of them can be expressed as a sequence of operations of type* (*a*). For if we add the rth row to the sth, then subtract the sth row of the resulting matrix from the rth, and finally add the rth row to the sth, the elements of the rth and sth rows (and the qth column) will be, successively:

$$(\epsilon_r^q,\ \epsilon_s^q); \qquad (\epsilon_r^q,\ \epsilon_r^q + \epsilon_s^q); \qquad (-\ \epsilon_s^q,\ \epsilon_r^q + \epsilon_s^q); \qquad (-\ \epsilon_s^q,\ \epsilon_r^q);$$

and the resulting matrix will thus be that obtained by changing the signs of the elements of the sth row and then interchanging the rth and sth rows.

In like manner, *the operation of interchanging two columns and changing the signs of the elements of one of them is expressible as a sequence of operations of type* (b).

5. In place of our two fundamental operations (a) and (b) we might have restricted ourselves to the operations:

(a') *To add a row to, or subtract it from, an adjacent row.*

(b') *To add a column to, or subtract it from, an adjacent column.*

We reduce the operation (a) to a sequence of operations (a') in the following manner. For definiteness, let us speak of the sth row as following the rth; in the reverse case "following" is to be replaced by "preceding" in the argument. Add each row from the rth to the $(s-1)$th inclusive to the next following row, beginning with the $(s-1)$th. Then, in the resulting matrix subtract each row from the $(r+1)$th to the $(s-1)$th inclusive from the following row, beginning at the $(r+1)$th. Next add each row from the $(r+1)$th to the $(s-2)$th inclusive to the following, beginning at the $(s-2)$th. Finally subtract each row from the rth to the $(s-2)$th from the following row, beginning at the rth.

If the reader will write out the expressions for the elements of the matrix in a single column, and the rows affected (from the rth to the sth), he will find that the resulting matrix only differs from our original matrix in having its rth row added to its sth (or subtracted from it). To perform the inverse operation we need only repeat the process, subtracting the rth row from the $(r+1)$th in the second step, and adding it to the $(r+1)$th in the last step; the other operations remaining as before.

As a similar argument holds for steps (b) and (b'), if we replace rows by columns, we conclude that *the transformations built up from steps* (a) *and* (b) *are no more general than those built up from steps* (a') *and* (b').

Determinant Factors.

6. Consider the set of γ-rowed $(0 < \gamma \leq \alpha,\ \gamma \leq \beta)$ determinants which can be formed from E by omitting $\alpha - \gamma$ of the rows and $\beta - \gamma$ of the columns in all possible ways. The highest common factor (H. C. F.) of such a set of determinants, if the determinants are not all zero, is denoted by D_γ and is called the γth *determinant factor** of E.

The determinant factors are unchanged when the matrix is operated on by transformations of type (a) *or* (b). For, consider the effect of an operation of type (a) which consists in adding the rth row to (or subtracting it from) the sth, on the γ-rowed determinants in question. All such de-

* Cf. Scott and Mathews' Determinants, p. 76.

terminants which do not contain elements from the sth row are obviously unaffected, while those that contain elements from both the rth row and the sth are not affected because of an elementary theorem on determinants. The remaining γ-rowed determinants, as A_γ, which contain elements from the sth row and not from the rth, are converted into determinants of the form $A_\gamma \pm A_\gamma'$ where A_γ' is the γ-rowed determinant obtained from A_γ by replacing the elements from the sth row by elements from the same columns of the matrix and from the rth row.

The proof for operations of type (b) is similar.

7. The following theorem has an application in Analysis Situs: *If a matrix E is such that each column either consists entirely of zeros or contains just two elements different from* 0, *one* $+1$ *and the other* -1, *all the determinant factors of the matrix are* $+1$ *or* -1.

The theorem follows immediately from the definition of a determinant factor, if we observe that any γ-rowed determinant formed by striking out $(\alpha - \gamma)$ rows and $(\beta - \gamma)$ columns of the given matrix has either two, none or one element in each column different from zero. If no column is of the third type the determinant is zero, since the sum of all the elements in each column is zero. If there is a column of the third type we evaluate the determinant with reference to such a column and then evaluate the minor with reference to a column with a single non-zero element in the minor, and so on. In this way we either arrive finally at ± 1 for the value of the determinant, or else come to a minor with two or no non-zero elements in each column, in which case the determinant is zero.

Reduction to Normal Form.

8. Let us now consider a series of reductions of the matrix E which can be effected by transformations of types (a) and (b). If the first column consists entirely of zeros, add one of the other columns to it. Thus by a transformation of type (b) E is converted into a matrix E_1 which has at least one non-zero element in the first column. If the first element of the first column is zero, add a row which contains a non-zero element in the first column to the first row. Thus by a transformation of type (a) E_1 is converted into a matrix E_2 for which the element of the first row and column is not zero.

We shall now prove that if this non-zero element ϵ_1^1 is not a factor of all the elements of the matrix, we can, by a series of transformations of types (a) and (b), replace it by a numerically smaller element different from zero.

First, if one of the elements in the first column, ϵ^1_r, is not divisible by ϵ_1^1, upon adding the first row to (or subtracting it from) the rth a number of

times equal to the largest integer in the quotient of ϵ^1_r by ϵ_1^1, an element numerically smaller than ϵ_1^1 is obtained in the first column and rth row. Then, on subtracting the rth row from (or adding it to) the first row the matrix is converted into one with a smaller non-zero element in place of ϵ_1^1. This has been done by a succession of operations of type (a). Similarly, if there were an element in the first row which did not contain ϵ_1^1 as a factor, transformations of type (b), strictly analogous to those of type (a) just described, could be set up which would reduce the numerical value of ϵ_1^1.

Second, if ϵ_1^1 is a factor of all the elements of the first row and first column, but is not a factor of the element in the rth row and sth column, ϵ_r^s, we proceed as follows. Upon subtracting the first column from (or adding it to) the sth $\epsilon_1^s/\epsilon_1^1$ times (transformations of type (b)), the first element in the sth column becomes zero, while the rth is still not divisible by ϵ_1^1, since it has been changed by a multiple of ϵ_1^1. If we now add the sth column to the first (an operation of type (b)), the element in the first row and column remains ϵ_1^1, while the rth element in the first column is now not divisible by ϵ_1^1. Hence we may replace ϵ_1^1 by a numerically smaller element by the method of the preceding paragraph.

If the element which replaces ϵ_1^1 is not a factor of all the elements of the matrix, it may be still further reduced by a repetition of the process described in the two paragraphs above. If this process be continued, we must arrive after a finite number of steps—the number being less than the absolute value of ϵ_1^1—at a matrix whose first element d_1 is a factor of all the others. When this point is reached, we may reduce all the elements in the first column except the first to zeros by operations of type (a), for we have merely to add the first row to (or subtract it from) any other row the number of times the first element of this row contains d_1. The elements of the first row, with the exception of the first, may be reduced to zeros by similar operations of type (b). It is evident that all the elements of the matrix thus obtained contain the first element as a factor.

Thus we arrive at a matrix E_3 in which the first element d_1 of the first column is the H. C. F. of all the elements of E_3 and in which all the other elements of the first row and of the first column are zero. By § 6, d_1 is the H. C. F. of all the elements of E.

9. Let $\overline{E}_3$ be the matrix obtained from E_3 by deleting its first row and first column. By § 8, $\overline{E}_3$ may be reduced to a matrix with a leading element which is the H. C. F. of all its elements, and having all the other elements of the first row and column zero.

As the transformations of types (a) and (b) which effect this reduction on $\overline{E}_3$ determine transformations of E_3 of the same type, which leave its

first row and first column unchanged, we may reduce the matrix E_3 to a matrix E_4 in which the first element of the main diagonal, d_1, is the H. C. F. of all the elements of the matrix, the second element of the main diagonal, d_2, is the H. C. F. of all the elements except d_1, and all the remaining elements of the first two rows and first two columns are zero.

By a continuation of this process we arrive by a finite sequence of operations of types (a) and (b) at a matrix:

$$(5) \qquad E^* = \begin{Vmatrix} d_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & d_r & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{Vmatrix},$$

in which all the elements are zero except a sequence of elements d_i, $(0 < i \leq r)$ common to the ith row and column and such that d_i is the H. C. F. of all d_j's such that $i \leq j \leq r$.

The d_i's may be positive or negative integers. We can make all except the last positive by a sequence of operations of type (a). For if d_i is negative, and we interchange the ith and rth rows, changing the sign of the elements in the rth, a permissible operation by § 4, and repeat the process, we arrive at a form in which d_i is positive and d_r has changed sign. We may thus obtain a form in which d_i $(i < r)$ is positive, and d_r will be positive or negative according as the number of negative signs in the form we started with was even or odd. We shall take this matrix, with at most one negative element, as the normal form E^* in the discussion which follows.

Each operation of type (a) amounts, according to § 3, to multiplying the matrix to which it is applied on the left by a square matrix of type A_0 of α rows, and each operation of type (b) amounts to multiplying the matrix to which it is applied on the right by a square matrix of type B_0 of β rows. Hence

$$(6) \qquad E^* = A \cdot E \cdot B,$$

where A is a product of matrices of type A_0 and B a product of matrices of type B_0. It is to be noted that the determinants of A and B are each $+1$.

Let us introduce the notation $D_1 = d_1$, $D_2 = d_1 \cdot d_2$, $\cdots$ $D_r = d_1 \cdot d_2 \cdot \cdots d_r$, and observe that D_γ $(0 < \gamma \leq r)$ is the H. C. F. of all the γ-rowed determinants which can be formed by striking out $\alpha - \gamma$ rows and $\beta - \gamma$ columns from E^*. That is, referring to § 6, they are the successive

determinant factors of E^*. Since E^* was derived from E by operations of types (a) and (b), they are also the determinant factors of E.

Since the D_i's are invariant under transformations of types (a) and (b), the d_i's, which are the quotients of successive D_i's, ($d_{i+1} = D_{i+1}/D_i$), are also invariant under these transformations. They are called *invariant factors* or elementary divisors.*

The number r is also invariant under transformations of types (a) and (b) and is called the *rank* of the matrix E.

The Matrices of Transformation.

10. In the special case where E is a square matrix of α rows whose determinant is $+1$, equation (6) implies that the determinant of E^* is $+1$. Hence $r = \alpha$, and the numbers d_i must be $+1$.

We therefore have:

$$A \cdot E \cdot B = I, \tag{7}$$

in which A is a product of matrices of type A_0, B a product of matrices of type B_0 and I is the identity matrix. We may write (7) in the form

$$E = A^{-1} \cdot I \cdot B^{-1} = A^{-1} \cdot B^{-1}. \tag{8}$$

It is evident from § 2 that when $\alpha = \beta$ every matrix of type B_0 can be regarded also as one of type A_0, and the same is true of matrices inverse to those of types A_0 or B_0. As the above equation shows that E is equal to a product of such matrices, we have the theorem: *Any square matrix of determinant unity is expressible as a product of matrices which may be considered to be of type A_0, or to be of type B_0.*

Hence to multiply a matrix E of α rows and β columns on the left by a square matrix of α rows and determinant unity is equivalent to operating on E by a sequence of operations of type (a); *and to multiply E on the right by a square matrix of β columns and determinant unity is equivalent to operating on E by a sequence of operations of type* (b).

Also, since we may write (8) in either of the forms:

$$B \cdot A \cdot E = I \qquad \text{or} \qquad E \cdot B \cdot A = I, \tag{9}$$

it follows that *if E is a square matrix of determinant unity, it may be reduced to the form I by operations on rows only, or by operations on columns only.*

11. In the case of a general matrix E, we have from (6)

$$A \cdot E = E^* \cdot B^{-1}. \tag{10}$$

Since the determinant of B^{-1} is 1, the H. C. F. of the elements of its first

* We shall use the term invariant factor, following Bôcher, Introduction to Higher Algebra, pp. 269–70, since the term elementary divisor is sometimes used in another sense.

row is 1. Hence the H. C. F. of the elements of the first row of the matrix $E^* \cdot B^{-1}$ is d_1. As a similar statement applies to the remaining rows, we have the theorem:

The matrix A has the property that the H. C. F. of the elements of the rth row of the matrix $A \cdot E$ is d_r, the rth invariant factor of E.

This suggests a method of building up A by means of the theorems:

(1) That for any set of integers ϵ_j^1 $(0 < j \leq \alpha)$, a set of integers a_1^j $(0 < j \leq \alpha)$ can be found which are relatively prime and such that

$$\sum_{j=1}^{j=\alpha} a_1^j \cdot \epsilon_j^1 = t_1$$

where t_1 is the H. C. F. of the α ϵ_j^1's; and

(2) That there exists a matrix A of determinant unity with the numbers a_1^j as the elements of its first row.

The derivation of equation (6) by this method is longer than that given in §§ 8 and 9 and is therefore omitted.

Diophantine Equations of the First Degree.

12. Consider the problem of finding the integral solutions of the following set of equations:

$$(11) \qquad \begin{aligned} \epsilon_1^1 x_1 + \epsilon_1^2 x_2 + \cdots + \epsilon_1^\beta x_\beta &= p_1, \\ \epsilon_2^1 x_1 + \epsilon_2^2 x_2 + \cdots + \epsilon_2^\beta x_\beta &= p_2, \\ \cdots\cdots\cdots & \\ \epsilon_\alpha^1 x_1 + \epsilon_\alpha^2 x_2 + \cdots + \epsilon_\alpha^\beta x_\beta &= p_\alpha. \end{aligned}$$

If X denotes the matrix of one column, and β rows

$$\left\| \begin{matrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_\beta \end{matrix} \right\|,$$

and P a similar matrix with $p_1, p_2 \cdots p_\alpha$ as the elements of its one column and α rows, equations (11) may be written

$$(12) \qquad E \cdot X = P.$$

But from (6) we have

$$(13) \qquad E = A^{-1} \cdot E^* \cdot B^{-1},$$

and consequently

$$(14) \qquad A^{-1} \cdot E^* \cdot B^{-1} \cdot X = P,$$

or

$$(15) \qquad E^* \cdot B^{-1} \cdot X = A \cdot P.$$

Let us set $Q = A \cdot P$, a matrix of one column and α rows, and denote its elements by $q_1, q_2 \cdots q_\alpha$. Also let $y_1, y_2 \cdots y_\beta$ be the elements of the matrix $Y = B^{-1} \cdot X$, which is of one column and β rows. Then (15) becomes:

$$E^* \cdot Y = Q, \tag{16}$$

which is equivalent to the set of α equations:

$$\begin{aligned} d_i y_i &= q_i \qquad (0 < i \leq r), \\ 0 &= q_j \qquad (r < j \leq \alpha), \end{aligned} \tag{17}$$

where r is the rank of E. If equations (17) are to be consistent, the q_j's must all be zero, and in this case the solution is:

$$\begin{aligned} y_i &= \frac{q_i}{d_i} && (0 < i \leq r), \\ y_j &\text{ is arbitrary} && (r < j \leq \beta). \end{aligned} \tag{18}$$

To express the condition that equations (17) be consistent and *solvable in integers*, in terms of the coefficients of (11), we proceed as follows. Form the "augmented matrix" of the system, a matrix $\overline{E}$ of α rows and $\beta + 1$ columns whose ith row has as its elements:

$$\epsilon_i^1, \epsilon_i^2, \cdots, \epsilon_i^\beta, -p_i.$$

The matrix $\overline{S}$ formed by multiplying $\overline{E}$ on the left by A will have as the elements of its ith row $(0 < i \leq \alpha)$:

$$s_i^1, s_i^2, \cdots, s_i^\beta, -q_i,$$

where the s_i^j's are the elements of the matrix:

$$S = \| s_i^j \| = A \cdot E.$$

Since multiplying the matrix S by B reduces it to the normal form, it may be reduced to the form E^* by operations on columns only; which shows that $\overline{S}$ may be reduced by operations on columns only to the form:

$$\overline{E}^* = \begin{Vmatrix} d_1 & 0 & \cdots & 0 & 0 & \cdots & 0 & -q_1 \\ 0 & d_2 & \cdots & 0 & 0 & \cdots & 0 & -q_2 \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & d_r & 0 & \cdots & 0 & -q_r \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & -q_\alpha \end{Vmatrix}. \tag{19}$$

In order that (11) be solvable at all, we found that q_i must be zero for

values of i greater than r. This shows that the rank of $\overline{E}^*$ is r. If in addition we require the solutions to be integers, q_i must be divisible by d_i for $i \leq r$. Hence $\overline{E}^*$ may be reduced to normal form by adding the ith column to the last q_i/d_i times. Hence its invariant factors must be the same as the d_i's, i.e., those of E. Conversely, if this condition is satisfied, each q_i will be divisible by the corresponding d_i, and the solutions of (11) will be integers.

Since $\overline{E}^*$ was obtained from $\overline{E}$ by elementary transformations, it has the same rank and invariant factors as $\overline{E}$. Hence we have proved the two theorems:

A necessary and sufficient condition that the equations (11) *have a set of integral solutions is that the augmented matrix* $\overline{E}$ *have the same rank and invariant factors as the matrix of the coefficients* E.

13. Since the solutions of (17) are given by (18), and since $X = B \cdot Y$, the solutions of (11) are:

$$x_i = \sum_{j=1}^{j=\beta} b_i^j y_j = \sum_{j=1}^{j=r} b_i^j \frac{q_j}{d_j} + \sum_{j=r+1}^{j=\beta} b_i^j y_j \qquad (0 < i \leq \beta), \tag{20}$$

in which $y_{r+1}, y_{r+2}, \cdots, y_\beta$ are arbitrary integers.

If the equations were homogeneous, the p_i's would all be zero, and hence the q_i's would also be zero. Hence the solutions would be of the form:

$$x_i = \sum_{j=r+1}^{j=\beta} b_i^j y_j \qquad (0 < i \leq \beta), \tag{21}$$

in which $y_{r+1}, y_{r+2}, \cdots, y_\beta$ are arbitrary integers.

Consequently, for such equations we have the theorem:

A set of linear homogeneous equations whose coefficients are integers has a set of $\beta - r$ *linearly independent solutions each of which is a set of relatively prime integers, if* β *is the number of unknowns and* r *the rank of the matrix of the coefficients. All other solutions in integers are linearly dependent on these* $\beta - r$ *linearly independent solutions, the coefficients of the linear relations being integers.*

This result was to be expected, since if a set of linear homogeneous equations are solvable in rational numbers, they are solvable in integers.

By comparing (20) and (21) we obtain the further result:

If one set of integers satisfying equations (11) *be given, the other solutions are obtained by adding to it the solutions of the homogeneous equations which result when the right members of* (11) *are replaced by zeros.*

The theorems of this paragraph were first given in complete form by H. J. S. Smith,* although he was anticipated to some extent by Heger.†

* Smith, H. J. S., On Systems of Linear Indeterminate Equations and Congruences, Philos. Transactions, Vol. 151, pp. 293 f. Collected Works, XII, pp. 367 ff.

† Heger, Ignaz, Mem. Vienna Academy, Vol. XIV, second part, p. 111.

Skew-Symmetric Matrices.

14. A skew-symmetric matrix is one in which

$$\epsilon_i^j = -\epsilon_j^i. \tag{22}$$

Let us define as the *conjugate* of a square matrix the matrix obtained from it by interchanging rows and columns. Evidently if a skew-symmetric matrix be pre-multiplied by any square matrix, and post-multiplied by the conjugate of this matrix, it will remain skew-symmetric.

If A is the matrix defined in § 9 such that

$$A \cdot E \cdot B = E^*, \tag{6}$$

the matrix $A \cdot E$, by § 11, has d_1 as the H. C. F. of the elements of the first row. Since multiplication on the right corresponds to operations on columns only and leaves the H. C. F. of the elements of the first row unchanged, d_1 is also the H. C. F. of the elements of the first row of $A \cdot E \cdot A'$ where A' denotes the conjugate of A. Since $A \cdot E \cdot A'$ is skew-symmetric, d_1 will also be the H. C. F. of the elements of its first column.

We reduce $A \cdot E \cdot A'$ further as follows: If the second element of the first row does not divide all the remaining elements in that row, let ϵ_1^j be one which it does not divide. Subtract the second column from (or add it to) the jth a number of times equal to the greatest integer in the quotient $\epsilon_1^j/\epsilon_1^2$, thus replacing ϵ_1^j by an element numerically less than ϵ_1^2. Upon subtracting the jth column from (or adding it to) the second, we obtain an element in the first place of the second column smaller than the one there before. All these operations leave the first column unchanged, and since the matrix was skew-symmetric, a similar set of operations on the rows reduces the matrix to a skew-symmetric matrix with the first element in the second column numerically smaller than before. By repeating these operations a sufficient number of times—at most $|\epsilon_1^2|$ times—this first element will be the highest common factor of the elements in the first row, and consequently of the elements of the matrix. When this condition is reached, we combine the second column with the other columns such a number of times that all the elements in the first row after the second will be zero, and perform similar operations on the rows. Then we combine the first column with the other columns such a number of times that all the elements in the second row after the first will be zero, and perform similar operations on the rows. This will reduce our matrix to the form:

$$E_1 = \begin{Vmatrix} 0 & d_1 & 0 & 0 & \cdots & 0 \\ -d_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \epsilon_3^4 & \cdots & \epsilon_3^a \\ 0 & 0 & \epsilon_4^3 & 0 & \cdots & \epsilon_4^a \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \epsilon_a^3 & \epsilon_a^4 & \cdots & 0 \end{Vmatrix}. \tag{23}$$

The matrix obtained from E_1 by deleting its first two rows and columns is skew-symmetric, and by applying the above process to it, in a way wholly analogous to the way by which we extended our initial process of reduction in § 9, we may, by a finite number of operations, reduce our matrix to one, $\bar{E} = \| \epsilon_i^j \|$, in which

$$\epsilon_{2i-1}^{2i} = d_i; \qquad \epsilon_{2i}^{2i-1} = -d_i \qquad (0 < i \leq p)$$

and the remaining elements are zero. That is, our matrix consists of a series of skew blocks of two non-zero elements each along the main diagonal, surrounded by zero elements.

Since in the above process we have always performed identical operations on rows and columns, we may write:

$$E = U \cdot \bar{E} \cdot U' \tag{24}$$

where U and U' are conjugate matrices whose determinants are $+1$.

Since interchanging the first and second, third and fourth, $\cdots$ $(2n - 1)$th and $2n$th rows, and changing the signs of the even rows would reduce this matrix to the usual normal form E^*, the d_i's appearing in $\bar{E}$ must be identical with those of E^*, i.e., the invariant factors of E. Hence, we have the result:

The invariant factors of a skew-symmetric matrix are equal in pairs, and the rank of such a matrix is an even number. A skew-symmetric matrix may be reduced to the "skew" normal form, $\bar{E}$, by multiplying on the left by a unimodular matrix U and on the right by its conjugate, U'.

Symmetric Matrices.

15. A *symmetric* matrix is one in which:

$$\epsilon_i^j = \epsilon_j^i. \tag{25}$$

Since a symmetric matrix retains its symmetry when we perform any operations on its rows, provided we perform the same operations on its columns, the question naturally arises whether a process similar to that of the preceding paragraph exists which will enable us to reduce such matrices to their normal form by means of a matrix and its conjugate. This question must be answered in the negative,* as is proved by the following example. The matrix

$$E = \begin{Vmatrix} 2 & 1 \\ 1 & 2 \end{Vmatrix}, \tag{26}$$

* On p. 189 of Scott and Mathews' Determinants the erroneous statement is made that symmetric matrices with integral elements can always be reduced to normal form by identical operations on rows and columns.

can not be reduced to its normal form,

$$E^* = \left\| \begin{matrix} 1 & 0 \\ 0 & 3 \end{matrix} \right\|, \tag{27}$$

by a matrix

$$U = \left\| \begin{matrix} a & b \\ c & d \end{matrix} \right\|, \tag{28}$$

and its conjugate U', where a, b, c and d are integers, since one of the conditions the elements of U would have to satisfy is:

$$2a^2 + 2ab + 2b^2 = 1. \tag{29}$$

The reduction and classification of symmetric matrices by identical operations on rows and columns is thus a problem of a different order from those which have been considered in this paper. Equivalence under such operations involves much more than the equivalence of invariant factors. This classification is the fundamental problem of the arithmetic theory of quadratic forms.*

Matrices with Elements Reduced, Modulo 2.

16. In many applications of matrices to Analysis Situs, it is found convenient to reduce the elements of the matrices modulo 2. On reducing modulo 2, the equation (6) becomes

$$\bar{A} \cdot \bar{E} \cdot \bar{B} = \bar{E}^*, \tag{30}$$

in which $\bar{E}$ can represent an arbitrary matrix of α rows and β columns whose elements are 0 and 1, $\bar{A}$ and $\bar{B}$ represent square matrices of determinant unity (mod. 2) of α and β rows respectively, and $\bar{E}$ is a matrix all of whose elements are 0 except a sequence of elements along the main diagonal which are 1. This follows from the fact that if one of the invariant factors of E^* is even, so are all the following invariant factors since they contain this one as a factor. The number of 1's in $\bar{E}^*$ is the rank of $\bar{E}^*$. It is less than or equal to the rank of E^*, and differs from it by the number of even invariant factors of E.

Symmetric Matrices, Modulo 2.

17. The theory of symmetric matrices, mod. 2, is not subject to the difficulties referred to in § 15. The reduction of such a matrix to normal form may be effected as follows: First interchange rows (performing the same interchange of columns) until the main diagonal consists of a series of 1's followed by a series of 0's. This can be effected by elementary

* Cf. Encyclopédie des Sciences Mathématiques, Tome I, vol. 3, p. 101.

transformations according to § 4 because the negative of any element is the same as the element itself, modulo 2. Add the first row to every row whose first element is a 1, and the first column to the corresponding columns. Repeat this for the second row and column performing a new interchange of rows and columns, if necessary, and continue until there are no elements different from zero in the main diagonal after those used.

The part of the matrix still to be normalized is now in the skew-symmetric form, since $+1 = -1$ (mod. 2) and may be normalized by the process of § 14. Thus by identical operations on rows and columns we have reduced our matrix to the form $\bar{E}^* = \| \epsilon_i^j \|$ in which:

$$\epsilon_i^i = 1\ (0 < i \leq p); \quad e_{p+2i-1}^{p+2i} = 1; \quad e_{p+2i}^{p+2i-1} = 1\ (0 < i \leq q),$$

and all the remaining elements of the matrix are zero. That is, the non-zero elements consist of a series of 1's in the main diagonal, followed by a series of skew blocks, each containing two 1's. If $p = 0$, this matrix can not be reduced further; but if $p \neq 0$, it may be reduced to a form containing one or two 1's in the main diagonal (according as p is odd or even) and a series of skew blocks, or to a form containing a series of 1's in the main diagonal and no skew blocks. This further reduction depends on the fact that a group of three 1's in the main diagonal of a matrix in the above form may be replaced by a single 1 in the main diagonal and a skew block of two.

The steps of the process in the case of a three-rowed square matrix are, first adding the first row and column to the second row and column respectively, then adding the third row and column to the first row and column respectively, and finally adding the second row and column to the third. The matrix becomes successively:

$$\begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{Vmatrix}, \quad \begin{Vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{Vmatrix}.$$

These steps may be made in reverse order to effect the inverse transformation, and are obviously typical of the steps which can be applied to any matrix $\bar{E}^*$ for which $p > 0$.

Symmetric Matrices, Modulo p.

18. In the case of an odd prime modulus, the reduction of a symmetric matrix is even simpler than in the modulo 2 case, and the normal form is a matrix all of whose elements are 0 except a sequence down the main diagonal. For, by a set of interchanges on rows, followed by similar ones on columns, we can obtain a non-zero element in the first row

of the matrix. If this is not the leading element, by adding the column containing it to the first column, and the corresponding row to the first row, we obtain a leading element, $\epsilon_1{}^1$, which is not zero. We may then add the first row to the others so as to make all the other elements in the first column 0, and operate similarly on the columns. The number of times, n, we must add the first row to a row with first element $\epsilon_r{}^1$ is given by solving the congruence:

$$(31) \qquad n\epsilon_1{}^1 + \epsilon_r{}^1 \equiv 0 \qquad (\text{mod. } p)$$

which has a root since p is prime. The reduction is continued as in § 9. By interchanges of rows and columns after we have reduced our matrix to a form similar to that given by (5), we can change the order of the d_i's, and have to decide on some definite order to get a single normal form. We may, for example, write the $p - 1$ non-zero elements of our system as the integers from 1 to $(p - 1)/2$ with plus and minus signs; and take as the normal order that of absolute value of the elements when written in this form.

AN ALGORISM FOR DIFFERENTIAL INVARIANT THEORY.

By Oliver E. Glenn.

It is my purpose to formulate in this paper, for the theory of differential invariants as derived by transformation of binary differential quantics, an algorism of fundamental simplicity* which I have described for algebraic concomitants in research papers written heretofore. Briefly stated the methods relate to certain irrational expressions in the arbitrary functions occurring in the coefficients of the transformations, which serve to define a domain of rationality R within which all differential invariants previously known are functions of certain elementary invariants, in R, and their derivatives, together with arbitrary functions and their derivatives. These elementary invariants, to be designated, in the present paper at least, as *invariant elements*, have served to unify and to simplify to an appreciable degree algebraic theories such as those of boolean† and orthogonal‡ concomitants, and in fact also that of the general algebraic concomitants, as, by their use, I developed a new proof of Gordan's theorem§ which is at least as simple as any other known proof of this important finiteness theorem.‖

1. Differential forms. Suppose that two quadratic differential forms

$$f = adx_1^2 + 2bdx_1dx_2 + cdx_2^2,$$
$$f' = Ady_1^2 + 2Bdy_1dy_2 + Cdy_2^2,$$

in which a, b, c are functions of x_1, x_2 and A, B, C are functions of y_1, y_2, can be so related by an arbitrary functional connection between the variables, that is,

$$x_1 = x_1(y_1, y_2), \qquad x_2 = x_2(y_1, y_2), \tag{1}$$

that when (1) is substituted in f it becomes f'. This implies transformation of the variables and also the differentials, the latter by the substitutions

$$T : dx_i = \frac{\partial x_i}{\partial y_1} dy_1 + \frac{\partial x_i}{\partial y_2} dy_2 \quad (i = 1, 2). \tag{2}$$

* Lemoine, "Considérations générales sur la mesure de la simplicité dans les sciences mathématiques, etc." Mathematical Papers, International Math. Congr., Chicago, 1893.

† Boole, Cambr. Math. Journ., vol. 3 (1843), p. 1.

‡ Elliott, Proc. Lond. Math. Soc., vol. 33 (1901), p. 226.

§ Cayley, Coll. Math. Papers, vol. 2, p. 250; Gordan, Journ. für Math., vol. 69 (1868), p. 323.

‖ O. E. Glenn, Trans. Amer. Math. Soc., vol. 20 (1919), p. 203.

Under these conditions

$$
(3)\qquad
\begin{aligned}
A &= a\left(\frac{\partial x_1}{\partial y_1}\right)^2 + 2b\,\frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_1} + c\left(\frac{\partial x_2}{\partial y_1}\right)^2,\\
B &= a\,\frac{\partial x_1}{\partial y_1}\frac{\partial x_1}{\partial y_2} + b\left(\frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} + \frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1}\right) + c\,\frac{\partial x_2}{\partial y_1}\frac{\partial x_2}{\partial y_2},\\
C &= a\left(\frac{\partial x_1}{\partial y_2}\right)^2 + 2b\,\frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_2} + c\left(\frac{\partial x_2}{\partial y_2}\right)^2.
\end{aligned}
$$

These three differential equations of the first order, if solved, would give the transformations (1).

We next write the general form F of order m under the notation

$$F = \sum_{r=0}^{m}\binom{m}{r} a_r dx_1^{m-r} dx_2^{r}$$

and express it as a symbolical mth power, employing equivalent symbols f, φ, $\cdots$, and writing $\partial f/\partial x_1 = f_1$, $\partial f/\partial x_2 = f_2$, f being, symbolically, a function of x_1, x_2. Thus we may write

$$F = (f_1 dx_1 + f_2 dx_2)^m = (df)^m = (d\varphi)^m = \cdots,$$

whence

$$(3_1)\qquad a_r = f_1^{m-r} f_2^{r} \quad (r = 0, \cdots, m).$$

2. **The domain** $R(1, T, \Delta)$. The poles of the transformations (2) in the differentials dx_1, dx_2 are the zeros of the linear forms

$$h_{\pm 1} = 2\,\frac{\partial x_2}{\partial y_1}\,dx_1 + \left(\frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_1} \pm \Delta\right) dx_2,$$

where

$$\Delta = \left[\left(\frac{\partial y_2}{\partial x_2} - \frac{\partial x_1}{\partial y_1}\right)^2 + 4\,\frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1}\right]^{\frac{1}{2}},$$

and two functions $f_{\pm 1}$ which satisfy the equalities

$$
(4)\qquad
\begin{aligned}
df_{+1} &= \frac{\partial f_{+1}}{\partial x_1}\,dx_1 + \frac{\partial f_{+1}}{\partial x_2}\,dx_2 \equiv h_{+1},\\
df_{-1} &= \frac{\partial f_{-1}}{\partial x_1}\,dx_1 + \frac{\partial f_{-1}}{\partial x_2}\,dx_2 \equiv h_{-1},
\end{aligned}
$$

will evidently be such that their functional determinant is equal to $4\Delta\,\partial x_2/\partial y_1$. Thus we may write $df_{\pm 1} = h_{\pm 1}$ provided $\Delta \neq 0$; the condition that the substitutions T be non-parabolic, and provided $\partial x_2/\partial y_1 \neq 0$, or that the functions x_1, x_2 be independent and contain both variables y_1, y_2 explicitly. Thus these considerations in connection with the poles of T give a transformation on the differentials whose coefficients appertain

to a domain $R(1, T, \Delta)$, the notation for which indicates that the functions therein are expressions with numerical coefficients in the functions, and partial derivatives thereof, occurring in the transformations (1) and (2), in the functional coefficients of F and their derivatives and so forth; and in Δ.

Besides this it is important to observe that integration of the equations $df_{\pm 1} = h_{\pm 1}$ would give x_1, x_2 as functions of f_{+1}, f_{-1} so that the sets x_1, x_2; y_1, y_2; f_{+1}, f_{-1} are functionally interrelated.

The quantics $df_{\pm 1}$ are formally invariantive under T; the multipliers in the invariant relations however are powers of the two factors, in the domain $R(1, T, \Delta)$, of

$$D = \frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1}.$$

That is, $df_{\pm 1}$ are differential covariants the invariant relations for which are

$$(5_1) \qquad df_{+1}' = \rho_{+1}^{-1} df_{+1}, \qquad df_{-1}' = \rho_{-1}^{-1} df_{-1} \equiv \rho_{+1} D^{-1} df_{-1},$$

primes indicating functions of y_1, y_2, and where

$$\rho_{\pm 1} = \frac{1}{2}\left(\frac{\partial x_1}{\partial y_1} + \frac{\partial x_2}{\partial y_2} \pm \Delta\right), \qquad \rho_{+1}\rho_{-1} = D.$$

3. The invariant elements. We employ henceforth the following abbreviations:

$$\frac{\partial x_1}{\partial y_1} = \alpha_1, \qquad \frac{\partial x_1}{\partial y_2} = \alpha_2, \qquad \frac{\partial x_2}{\partial y_1} = \beta_0, \qquad \frac{\partial x_2}{\partial y_2} = \beta_1,$$

whence the inverse of the transformations (4) takes the form

$$T': \begin{cases} dx_1 = (-4\beta_0\Delta)^{-1}[(\gamma_1 - \Delta)df_{+1} - (\gamma_1 + \Delta)df_{-1}], \\ dx_2 = (-4\beta_0\Delta)^{-1}[2\beta_0(-df_{+1} + df_{-1})] \quad (\gamma_1 = \beta_1 - \alpha_1). \end{cases}$$

The substitution T' operated upon the form F gives a unique expansion of the latter whose symbolical expression is

$$(5) \quad F = (-4\beta_0\Delta)^{-m}\{[(\gamma_1 - \Delta)f_1 - 2\beta_0 f_2]df_{+1} + [-(\gamma_1 + \Delta)f_1 + 2\beta_0 f_2]df_{-1}\}^m.$$

Hence

$$(6) \qquad F = \sum_{i=0}^{m} \binom{m}{i} \varphi_{m-2i}\, df_{+1}^{m-i} df_{-1, i},$$

in which

$$(7) \quad \varphi_{m-2i} = [(\gamma_1 - \Delta)f_1 - 2\beta_0 f_2]^{m-i}[-(\gamma_1 + \Delta)f_1 + 2\beta_0 f_2]^i \times (-4\beta_0\Delta)^{-m} \quad (i = 0, \cdots, m).$$

THEOREM I. *The functions φ_{m-2i}, which belong to the domain $R(1, T, \Delta)$, are differential invariants.*

Transformation of F by (1), (2) gives $F' = F$, and expansion of F in the arguments df_{+1}, df_{-1} yields the formula (6). We can also expand F' in the covariant arguments

$$df_{\pm 1}' = 2\frac{\partial x_2}{\partial y_1}dy_1 + \left(\frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_1} \pm \Delta\right)dy_2,$$

with the result

$$F' = \sum_{i=0}^{m}\binom{m}{i}\varphi_{m-2i}'df_{+1}'^{m-i}df_{-1}'^{i}. \tag{8}$$

Hence, by substitution from (5_1), we obtain, after equating coefficients of like powers of $df_{\pm 1}$, the following invariant relations for φ_{m-2i}:

$$\varphi_{m-2i}' = \rho_{+1}^{m-2i}D^i\varphi_{m-2i} \quad (i = 0, \cdots, m). \tag{9}$$

We can prove, accordingly, a quite general theorem on the reducibility of a differential invariant. Transformation T' can be written

$$T': \begin{cases} dx_1 = \dfrac{\partial x_1}{\partial f_{+1}}df_{+1} + \dfrac{\partial x_1}{\partial f_{-1}}df_{-1}, \\[2ex] dx_2 = \dfrac{\partial x_2}{\partial f_{+1}}df_{+1} + \dfrac{\partial x_2}{\partial f_{-1}}df_{-1}, \end{cases}$$

and is as general as the transformation T. Let B be any differential invariant whatsoever of F under T for which $B' = \alpha B$, that is,

$$B(a_r'; dy_1, dy_2) = \alpha B(a_r; dx_1, dx_2). \tag{10}$$

Then a cognate relation holds when the transformed form is (6), viz.,

$$B(\varphi_{m-2r}; df_{+1}, df_{-1}) = \alpha B(a_r; dx_1, dx_2).$$

It is desirable however to state this result in more general terms. Note first that any arbitrary function u of a_r $(r = 0, \cdots, m)$, dx_1, dx_2, by virtue merely of its being arbitrary, satisfies the relation $u' = u$ under the transformation T. Let Ω be a function of the functions a_r and of their x_1, x_2 derivatives and dx_1, dx_2, and also of a certain number of other specified functions $u_1, u_2, \cdots$ for which a relation $u' = u$ holds, among which functions some which are arbitrary functions may be comprised, together with the x_1, x_2 derivatives of $u_1, u_2, \cdots$. Let Ω' be the same function of a_r' and the y_1, y_2 derivatives of a_r' $(r = 0, \cdots, m)$ and of dy_1, dy_2, together with $u_1', u_2', \cdots$ and their y_1, y_2 derivatives. Then, if $\Omega' = \alpha\Omega$, Ω is called a differential parameter. In particular when there are no arbitrary functions actually involved in Ω it is a differential invariant. Hence,

THEOREM II. *Every differential parameter Ω is reducible in the domain*

$R(1, T, \Delta)$, in terms of the $m + 1$ invariant elements φ_{m-2i} $(i = 0, \cdots, m)$ and their x_1, x_2 derivatives, together with arbitrary functions and the differential covariants df_{+1}, df_{-1}.

Note that all of the elements in terms of which Ω is thus reducible are invariantive with the exception of the derivatives of φ_{m-2i} $(i = 0, \cdots, m)$, and we shall prove that these derivatives, also, satisfy an extended form of invariant relation.

The exact form in which an invariant appears as a function of invariant elements will be illustrated by means of two well-known differential parameters, viz.,

$$\begin{aligned}
\Delta_1 u &= a_2 \frac{\partial u}{\partial x_1}\frac{\partial u}{\partial x_1} + 2a_1 \frac{\partial u}{\partial x_1}\frac{\partial u}{\partial x_2} + a_0 \frac{\partial u}{\partial x_2}\frac{\partial u}{\partial x_2} \\
&= \varphi_{-2}\left(\frac{\partial u}{\partial f_{+1}}\right)^2 + 2\varphi_0 \frac{\partial u}{\partial f_{+1}}\frac{\partial u}{\partial f_{-1}} + \varphi_2\left(\frac{\partial u}{\partial f_{-1}}\right)^2, \\
\nabla(u, v) &= a_2 \frac{\partial u}{\partial x_1}\frac{\partial v}{\partial x_1} + a_1\left(\frac{\partial u}{\partial x_1}\frac{\partial v}{\partial x_2} + \frac{\partial u}{\partial x_2}\frac{\partial v}{\partial x_1}\right) + a_0 \frac{\partial u}{\partial x_2}\frac{\partial v}{\partial x_2} \\
&= \varphi_{-2}\frac{\partial u}{\partial f_{+1}}\frac{\partial v}{\partial f_{+1}} + \varphi_0\left(\frac{\partial u}{\partial f_{+1}}\frac{\partial v}{\partial f_{-1}} + \frac{\partial u}{\partial f_{-1}}\frac{\partial v}{\partial f_{+1}}\right) + \varphi_2 \frac{\partial u}{\partial f_{-1}}\frac{\partial v}{\partial f_{-1}}.
\end{aligned}$$

These appertain to the case $m = 2$. A pure invariant for this case is the discriminant $\delta = a_1^2 - a_0 a_2$, and

$$\delta = 4\beta_0^2\Delta^2(\varphi_0^2 - 4\varphi_2\varphi_{-2}).$$

4. Types of parameters. The algorism established by the preceding theorems affords a classification of differential parameters into types some of which consist of invariants of what seem to be entirely new categories.

(*a*) *The formal type.* The quantic F is formally analogous to a binary algebraical quantic, the transformations (2) to the linear transformations of such a quantic employed in algebraic invariant theory and the relations (3) and their generalizations to the transformations of the induced group in that theory. There is therefore a type of concomitant and a theory appertaining thereto closely analogous to formal algebraic invariant theory, the concomitants being expressible rationally in terms of symbols* $f_1, f_2, \varphi_1, \varphi_2, \cdots$; dx_1, dx_2, and reducible in terms of invariant elements $\varphi_m, \cdots, \varphi_{-m}$; df_{+1}, df_{-1} in the manner illustrated in the example of the discriminant of the quadratic in the preceding section. There are differential identities for the reduction of such concomitants which can be cast in a formal mould so as to be, in effect, similar to the symbolism of

* Maschke, Trans. Amer. Math. Soc., vol. 1 (1900), p. 197, and ibid., vol. 4 (1903), p. 445. Haskins, Trans. Amer. Math. Soc., vol. 3 (1902), p. 71. A. W. Smith, Trans. Amer. Math. Soc., vol. 7 (1906), p. 33. Ricci and Levi-Civita, Math. Annalen, vol. 54 (1901).

Aronhold and Clebsch. Thus when we abbreviate as follows:

$$\frac{\partial U}{\partial x_i} = U_i, \qquad \frac{\partial^2 U}{\partial x_i \partial x_k} = U_{ik},$$

$$U_1 V_2 - U_2 V_1 = (U,\ V), \tag{11}$$

and assume $m = 2$, we get

$$\alpha'(F,\ \varphi)' = \alpha(F,\ \varphi) \quad (\alpha = 1/\sqrt{\delta}),$$

where F, φ are any parameters, and

$$\Delta_1 u = \alpha^2 (F,\ u)^2,$$
$$\nabla(u,\ v) = \alpha^2 (F,\ u)(F,\ v),$$
$$(a,\ b)(c,\ d) + (a,\ c)(d,\ b) + (a,\ d)(b,\ c) = 0.$$

(*b*) *The extended formal type.* From (3_1) there follows

$$\frac{\partial a_r}{\partial x_1} = ((m-r) f_{11} f_2 + r f_{21} f_1) f_1^{m-r-1} f_2^{r-1},$$
$$\frac{\partial a_r}{\partial x_2} = ((m-r) f_{12} f_2 + r f_{22} f_1) f_1^{m-r-1} f_2^{r-1} \quad (r = 0,\ \cdots,\ m).$$

Hence differential parameters involving the first derivatives of the functions a_r are represented symbolically by means of expressions constructed from the combinations below and generalizations to higher derivatives will be obvious:

$$f_1^{m-r-1} f_2^r f_{11}, \qquad \varphi_1^{m-r-1} \varphi_2^r \varphi_{11}, \qquad f_1^{m-r} f_2^{r-1} f_{21}, \qquad f_1^{m-r-1} f_2^r f_{12},$$

and so forth.

An example* is the following for the quadratic form which we write under the notation

$$F = \sum_{i,k=1}^{2} a_{ik} dx_i dx_k \quad (a_{ki} = a_{ik}):$$

$$\Delta_2 u = \alpha(F,\ \alpha(F,\ u)) = \sum_{r,s} A_{rs} \frac{\partial^2 u}{\partial x_r \partial x_s} - \sum_{r,s,i,k} A_{rs} A_{ik} \begin{bmatrix} rs \\ k \end{bmatrix} \frac{\partial u}{\partial x_k},$$

where $\begin{bmatrix} rs \\ k \end{bmatrix}$ is the so-called triple index symbol due to Christoffel;

$$\begin{bmatrix} rs \\ k \end{bmatrix} = \frac{1}{2} \left(\frac{\partial a_{rk}}{\partial x_s} + \frac{\partial a_{sk}}{\partial x_r} - \frac{\partial a_{rs}}{\partial x_k} \right) = f_k f_{rs},$$

and A_{rs} denotes the minor of a_{rs} in $\delta = |a_{rs}|$. The expression in terms of the symbols is

$$\Delta_2 u = \alpha^2 [f_1 (f_{12} u_2 - f_{22} u_1) - f_2 (f_{11} u_2 - f_{12} u_1)] + \cdots.$$

The symbolism of the paragraphs (*a*), (*b*) was discovered by Maschke. Expression of $\Delta_2 u$ in terms of invariant elements is obtained by forming its invariant relation, according to theorem II.

* Beltrami's second differential parameter. Compare J. E. Wright, Invariants of Quadratic Differential Forms (Cambridge Tracts, 1908).

(*c*) *The orthogonal type.* I shall designate as the orthogonal type of differential parameter those which can be generated in totality by forming rational expressions in invariant elements φ_{m-2i} $(i = 0, \cdots, m)$, df_{+1}, df_{-1} which simplify by multiplication into functions appertaining to the domain $R(1, T, 0)$. The essential forms from which to construct this totality are evidently $P \pm Q$ where P is of the type

$$P = \varphi_m^{x_0}\varphi_{m-2}^{x_1}\cdots\varphi_{-m}^{x_m}df_{+1}^{\sigma_1}df_{-1}^{\sigma_2}, \tag{12}$$

and Q is the conjugate of P,

$$Q = \varphi_{-m}^{x_0}\varphi_{-(m-2)}^{x_1}\cdots\varphi_m^{x_m}df_{-1}^{\sigma_1}df_{+1}^{\sigma_2}. \tag{13}$$

An example for the quadratic quantic, $F = a_0dx_1^2 + 2a_1dx_1dx_2 + a_2dx_2^2$, is the following:

$$\begin{aligned}\Delta(\varphi_2 df_{-1}^2 - \varphi_{-2}df_{-1}^2) &= \left[a_0\frac{\partial x_1}{\partial y_1} - a_0\frac{\partial x_2}{\partial y_2} + 2a_1\frac{\partial x_2}{\partial y_1}\right]dx_1^2 + \left[2a_0\frac{\partial x_1}{\partial y_2} + 2a_2\frac{\partial x_2}{\partial y_1}\right]dx_1dx_2 \\ &\quad + \left[2a_1\frac{\partial x_1}{\partial y_2} + a_2\frac{\partial x_2}{\partial y_2} - a_2\frac{\partial x_1}{\partial y_1}\right]dx_2^2.\end{aligned}$$

(*d*) *The extended orthogonal type.* The x_1, x_2 derivatives of an invariant element φ_{m-2i} are not invariantive. In the quadratic case for instance, or generally, we can obtain the relations which replace invariant relations for $\partial\varphi_{m-2i}/\partial x_j$ $(j = 1, 2)$ by applying to the members of the relation (9) the operators

$$\frac{\partial}{\partial y_1} = \frac{\partial x_1}{\partial y_1}\frac{\partial}{\partial x_1} + \frac{\partial x_2}{\partial y_1}\frac{\partial}{\partial x_2}, \qquad \frac{\partial}{\partial y_2} = \frac{\partial x_1}{\partial y_2}\frac{\partial}{\partial x_1} + \frac{\partial x_2}{\partial y_2}\frac{\partial}{\partial x_2}. \tag{14}$$

Notwithstanding this fact certain rational combinations of invariant elements, arbitrary functions, *and their derivatives* will belong to the domain $R(1, T, 0)$ and be invariantive. Such expressions will be called differential parameters of the extended orthogonal type.

(*e*) *Unclassified parameters.* Certain invariants, such as some irrational expressions in invariants of the above types, will not belong to any of the above categories and such as do not we leave unclassified.

5. Parameters of the orthogonal type. The invariant systems determined in this section are derived by means of the irreducible solutions of a certain linear diophantine equation and the generality of the conceptions involved is such that the same equation and the analogous algorism concerning invariant elements yields the corresponding systems for binary algebraic forms under orthogonal substitutions (the orthogonal invariants proper), under boolean transformations, the transformations of Einstein (invariants of relativity) and the general systems in $R(1, T, 0)$, all of which I have treated previously having published enumerations

for the orders from one to five inclusive. Hence I treat in detail in this paper the parameters of the orthogonal type for the differential quantic of order six.

We are concerned with concomitant expressions in the invariant elements which belong to $R(1, T, 0)$.

To construct such an expression from products such as P in (12) it is necessary, though not sufficient, that the exponent of ρ_{+1} in the invariant relation for P:

$$P' = \rho_{+1}{}^a D^b P, \tag{15}$$

should be zero (compare (9)). That is,

$$a = \sum_{i=0}^{m} (m - 2i)x_i - \sigma_1 + \sigma_2 = 0. \tag{16}$$

The concomitants P, Q do not belong to $R(1, T, 0)$, but they are in correspondence with $P + Q$, $P - Q$ which, when deprived of irrelevant factors, do appertain to that domain. Hence, according to Hilbert's lemma, the finite set of irreducible solutions of (16) furnish the exponents $x_0, \cdots, x_m, \sigma_1, \sigma_2$ of a complete system in the domain $R(1, T, 0)$. The system for the quantic

$$F = \sum_{r=0}^{6} \binom{6}{r} a_r dx_1{}^{6-r} dx_2{}^r$$

is obtained, therefore, by solving the equation

$$6x_0 + 4x_1 + 2x_2 + \sigma_2 = \sigma_1 + 2x_4 + 4x_5 + 6x_6, \tag{17}$$

and if an irreducible solution giving a product P is $(x_0, x_1, x_2, x_4, x_5, x_6, \sigma_1, \sigma_2)$ the solution which gives the conjugate product Q is $(x_6, x_5, x_4, x_2, x_1, x_0, \sigma_2, \sigma_1)$. The table below furnishes the irreducible solutions, in conjugate pairs, under the suggestive notation $\alpha \pm_1 = P \pm Q$. Thus, for example, a quadratic covariant is

$$\lambda_{\pm 1} = \varphi_6 \varphi_{-4}{}^2 df_{-1}{}^2 \pm \varphi_{-6} \varphi_4{}^2 df_{+1}{}^2.$$

To solve (17) I wrote it in the form

$$6x + 4y + 2z + w = 0, \tag{18}$$

$$x = x_0 - x_6, \qquad y = x_1 - x_5, \qquad z = x_2 - x_4, \qquad w = \sigma_2 - \sigma_1. \tag{19}$$

An appropriate number of solutions of (18) in both positive and negative integers being determined, the sets for (17) are furnished by solving (19). Thus the problem is subdivided into a large number of mutually exclusive subproblems. Including the invariant φ_0 not furnished by (17) and the covariant $\eta = df_{+1}df_{-1}$, the number of concomitants in the system is 31. There are 14 invariants ($\sigma_1 = \sigma_2 = 0$) and 17 covariants.

	x_0	x_1	x_2	x_4	x_5	x_6	σ_1	σ_3
α	1					1		
β		1			1			
γ			1	1				
$\delta \pm 1$		1		2				
			2		1			
$\epsilon \pm 1$	1			3				
			3			1		
$\zeta \pm 1$	1			1	1			
		1	1			1		
$\theta \pm 1$	1		1		2			
		2		1		1		
$\iota \pm 1$	2				3			
		3				2		
$\kappa \pm 1$			1				2	
				1				2
$\lambda \pm 1$		2				1	2	
	1				2			2
$\mu \pm 1$	1				1		2	
		1				1		2
$\nu \pm 1$	1			2			2	
			2			1		2
$\varphi \pm 1$	1			1			4	
			1			1		4
$\rho \pm 1$		1					4	
					1			4
$\sigma \pm 1$		1		1			2	
			1		1			2
$\tau \pm 1$	1						6	
						1		6

6. Parameters of the extended orthogonal type. From (8) it follows, by correspondence, that if $B(a_r;\, dx_1,\, dx_2,\, \cdots)$ is any differential parameter of F there exists a relation

$$B(\varphi_{m-2r}';\, df_{+1}',\, df_{-1}',\, \cdots) = QB(\varphi_{m-2r};\, df_{+1},\, df_{-1},\, \cdots).$$

We next prove certain results concerning the derivatives of invariant elements. From equations (14),

$$\frac{\partial}{\partial y_1} = \alpha_1 \frac{\partial}{\partial x_1} + \beta_0 \frac{\partial}{\partial x_2}, \qquad \frac{\partial}{\partial y_2} = \alpha_2 \frac{\partial}{\partial x_1} + \beta_1 \frac{\partial}{\partial x_2}, \tag{20}$$

and hence one obtains, by differentiation of the relation

$$\varphi_{m-2i}' = \rho_{+1}{}^{m-2i} D^i \varphi_{m-2i} = Q^{(i)} \varphi_{m-2i} \quad (i = 0, \cdots, m), \tag{21}$$

the following formulas:

$$\frac{\partial^r \varphi_{m-2i}'}{\partial y_1{}^{r-s} \partial y_2{}^s} = \alpha_1{}^{r-s} \beta_1{}^s Q^{(i)} \frac{\partial^r \varphi_{m-2i}}{\partial x_1{}^{r-s} \partial x_2{}^s} + \delta_r{}^{(i)} \tag{22}$$

$$(r = 0, 1, 2, \cdots; \quad s = 0, 1, \cdots, r; \quad i = 0, 1, \cdots, m),$$

where

$$\delta_r{}^{(i)} = \frac{\partial^{r-1} \varphi_{m-2i}}{\partial x_1{}^{r-s-1} \partial x_2{}^s} \left(\alpha_1 \frac{\partial}{\partial x_1} + \beta_0 \frac{\partial}{\partial x_2} \right) \left(\alpha_1{}^{r-s-1} \beta_1{}^s Q^{(i)} + \delta_{r-1}{}^{(i)} \right) + \alpha_1{}^{r-s-1} \beta_0 \beta_1{}^s Q^{(i)} \frac{\partial^r \varphi_{m-2i}}{\partial x_1{}^{r-s-1} \partial x_2{}^{s+1}}.$$

Thus while the derivatives of the invariant elements are not invariantive they satisfy equations which are obtained by adding increments $\delta_r{}^{(i)}$ to what would be, except for the increments, invariant relations of regular type for these derivatives. Hence we obtain

THEOREM III. *In order that a homogeneous function of the coefficients a_r and their derivatives,*

$$B(a_r;\ \partial^t a_r / \partial x_1{}^{t-u} \partial x_2{}^u;\ dx_1,\ dx_2),$$

should be a concomitant, it is necessary that the increment to

$$B(\varphi_{m-2r};\ \partial^t \varphi_{m-2r} / \partial x_1{}^{t-u} \partial x_2{}^u;\ df_{+1} df_{-1}), \tag{23}$$

under the appropriate equations (22), *should be zero. Subject to evident conditions of isobarism, this condition is also sufficient.*

When $r = 1$ we derive from the relations (22), which then reduce to

$$\frac{\partial \varphi_{m-2i}'}{\partial y_2} = \alpha_2 Q^{(i)} \frac{\partial \varphi_{m-2i}}{\partial x_1} + \beta_1 Q^{(i)} \frac{\partial \varphi_{m-2i}}{\partial x_2} + (\alpha_2 Q_{x_1}{}^{(i)} + \beta_1 Q_{x_2}{}^{(i)}) \varphi_{m-2i},$$

$$\frac{\partial \varphi_{m-2i}'}{\partial y_1} = \alpha_1 Q^{(i)} \frac{\partial \varphi_{m-2i}}{\partial x_1} + \beta_0 Q^{(i)} \frac{\partial \varphi_{m-2i}}{\partial x_2} + (\alpha_1 Q_{x_1}{}^{(i)} + \beta_0 Q_{x_2}{}^{(i)}) \varphi_{m-2i},$$

the formulae of transvection (cf. (11)),

$$\begin{aligned} \psi_{m-2i}' &= (\varphi_{m-2i}', Q^{(i)\prime}) = Q^{(i)} D(\varphi_{m-2i}, Q^{(i)}) = \rho_{+1}{}^{m-2i} D^{i+1}(\varphi_{m-2i}, Q^{(i)}) \\ &= Q^{(i)} D \psi_{m-2i} = Q_1{}^{(i)} \psi_{m-2i} \quad (i = 0, \cdots, m), \end{aligned} \tag{24}$$

and this establishes the following theorem:

THEOREM IV. *The transvectants ψ_{m-2i}, which belong to the domain $R(1, T, \Delta)$, are relative differential parameters involving first order derivatives of invariant elements φ_{m-2i} $(i = 0, \cdots, m)$, and consequently also first order derivatives of coefficients a_r of the ground form F.*

The relation (24) expressing the invariancy of ψ_{m-2i} is formally identical, save for the replacement of i by $i + 1$ in the exponent of D, with the corresponding relation for φ_{m-2i}. We can therefore form relations in the derivatives of the parameters ψ_{m-2i} similar to those employed in the derivation of the transvectants (24) and thus describe a process of iteration whose equivalent, by formula, is

$$((\varphi_{m-2i}', Q^{(i)'}), Q_1^{(i)'}) = \rho_{+1}^{m-2i} D^{i+2}((\varphi_{m-2i}, Q^{(i)}), Q_1^{(i)}). \tag{25}$$

Thus

$$\psi_{m-2i}^{(2)} = ((\varphi_{m-2i}, Q^{(i)}), Q^{(i)})$$

is a differential parameter involving derivatives of the second degree of invariant elements, and the extension to parameters of the rth iteration, functions of the rth derivatives of the invariant elements, is evident. Upon the basis of these iterated transvectants we construct a theory of parameters of the extended orthogonal type.

The invariant $\psi_{-(m-2i)}$ is conjugate to ψ_{m-2i}; hence systems of parameters in $R(1, T, 0)$, of the extended orthogonal type, involving, as to derivatives, powers of first derivatives, only, of invariant elements, can be formed upon the basis of combinations $P \pm Q$, where

$$P = \varphi_m^{x_0} \varphi_{m-2}^{x_1} \cdots \varphi_{-m}^{x_m} \psi_m^{y_0} \psi_{m-2}^{y_1} \cdots \psi_{-m}^{y_m} df_{+1}^{\sigma_1} df_{-1}^{\sigma_2}$$

and Q is the product conjugate to P. That is, in accordance with the lemma of Hilbert, a complete system is given by forms $\alpha_{\pm 1} = P \pm Q$ obtained from the totality of sets of irreducible solutions of the linear diophantine equation

$$\sum_{r=0}^{m} (m - 2r) x_r + \sum_{s=0}^{m} (m - 2s) y_s - \sigma_1 + \sigma_2 = 0. \tag{26}$$

The complete system for the differential quadratic is obtained, therefore, from the irreducible solutions of the equation

$$2x_0 + 2y_0 + \sigma_2 = 2x_2 + 2y_2 + \sigma_1,$$

shown in the table below:

	x_0	x_2	y_0	y_2	σ_1	σ_2
n					1	1
o	1	1				
p			1	1		
$q_{\pm 1}$	1			1		
		1	1			
$r_{\pm 1}$			1		2	
				1		2
$s_{\pm 1}$		1				2
	1				2	

The system consists of 11 forms, viz., φ_0, ψ_0 and

$$
\begin{aligned}
o &= \varphi_2\varphi_{-2}, \qquad p = (\varphi_2, \rho_{+1}{}^2)(\varphi_{-2}, \rho_{-1}{}^2),\\
q_{\pm 1} &= \varphi_2(\varphi_{-2}, \rho_{-1}{}^2) \pm \varphi_{-2}(\varphi_2, \rho_{+1}{}^2),\\
r_{\pm 1} &= (\varphi_2, \rho_{+1}{}^2)df_{+1}{}^2 \pm (\varphi_{-2}, \rho_{-1}{}^2)df_{-1}{}^2,\\
s_{\pm 1} &= \varphi_2 df_{+1}{}^2 \pm \varphi_{-2}df_{-1}{}^2, \qquad \eta = df_{+1}df_{-1}.
\end{aligned} \tag{27}
$$

Note the equality of the indices, i.e., the power of D appearing as a multiplier in the invariant relations for the terms of each binomial quantic in the list. Furthermore, the corresponding index for P is equal to that for Q but we can prove a still more general result. Let the first iteration of the transvectant $\psi_{m-2i} = \psi_{m-2i}{}^{(1)}$ in relation (25) be designated by $\psi_{m-2i}{}^{(2)}$:

$$\psi_{m-2i}{}^{(2)} = ((\varphi_{m-2i}, Q^{(i)}), Q_1{}^{(i)}) \quad (i = 0, \cdots, m),$$

and let the kth iterated transvectant be $\psi_{m-2i}{}^{(k+1)}$ so that $\varphi_{m-2i} = \psi_{m-2i}{}^{(0)}$. Then the conjugate to $\psi_{m-2i}{}^{(k+1)}$ is $\psi_{-(m-2i)}{}^{(k+1)}$ and the concomitant Q which is conjugate to

$$P = \prod_{i=0}^{m}\prod_{k=0}^{n} \psi_{m-2i}{}^{(k)x_{ik}} df_{+1}{}^{\sigma_1} df_{-1}{}^{\sigma_2} \tag{28}$$

is obtained by the corresponding changes of sign of subscripts.

THEOREM V. *The quantics $\alpha_{\pm 1} = P \pm Q$ are differential invariants of the extended orthogonal type appertaining to the domain $R(1, T, 0)$. They involve derivatives of invariant elements, and therefore of the functional coefficients a_r of the ground-form F itself, of all orders from zero to the nth and constitute an infinitude of quantics which possesses the property of finiteness. A complete system is given by the finite set of irreducible solutions in positive integers of the linear diophantine equation*

$$a = \sum_{k=0}^{n} \sum_{i=0}^{m} (m - 2i) x_{ik} + \sigma_2 - \sigma_1 = 0. \tag{29}$$

We need only the proof that the index for P equals that of Q. These indices are, respectively,

$$\begin{aligned} \alpha &= \sum_{k=0}^{n} \sum_{i=0}^{m} (i + k) x_{ik} - \sigma_2, \\ \beta &= \sum_{k=0}^{n} \sum_{i=0}^{m} (m + k - i) x_{ik} - \sigma_1. \end{aligned} \tag{30}$$

But $\beta - \alpha = a = 0$; hence $\alpha = \beta$.

I remark in conclusion that properly chosen polynomials in the expressions characteristic of parameters of the extended orthogonal type, no assumption here being made that these are generally expressible in any particular form except as polynomials in invariant elements and the derivatives of these, will be, when multiplied out, expressions which are free from the functions involved in the transformations, that is, parameters which belong to the domain $R(1, 0, 0)$. These are the concomitants to which the investigations of former writers relate. It is apparent that a finiteness theorem can be stated for the parameters in the latter domain and treated in the way analogous to that exemplified in my proof of the theorem of Gordan, quoted previously in this paper, and these and other developments would probably be of sufficient importance to warrant detailed treatment.

Differential parameters of the orthogonal types, containing arbitrary functions other than those involved in the transformations (1) and (2), are obtained from the relations

$$D \frac{\partial}{\partial x_2} = \alpha_1 \frac{\partial}{\partial y_2} + \alpha_2 \left(- \frac{\partial}{\partial y_1} \right), \qquad D \left(- \frac{\partial}{\partial x_1} \right) = \beta_0 \frac{\partial}{\partial y_2} + \beta_1 \left(- \frac{\partial}{\partial y_1} \right). \tag{31}$$

If $B(a_r; \partial^s a_r / \partial x_1^{s-t} \partial x_2^t; dx_1, dx_2)$ is any covariantive parameter, then, a parameter involving an additional arbitrary function u is

$$B(a_r; \partial^s a_r / \partial x_1^{s-t} \partial x_2^t; \partial / \partial x_2, - \partial / \partial x_1) u.$$

This method can be employed to derive an infinitude of concomitants involving arbitrary functions $u, v, \cdots$, from the systems derived in preceding sections.

THE UNIVERSITY OF PENNSYLVANIA,
July, 1920.

THE GENERAL THEORY OF CYCLIC-HARMONIC CURVES.

By Robert E. Moritz.

1. Introduction.

1.1. *Cyclic-harmonic motion* may appropriately be defined as motion resulting from the composition of simple-harmonic motion in a straight line with uniform rotatory motion about a fixed point in this line. The locus of the resultant motion is a *cyclic-harmonic curve.*

1.2. Let the simple-harmonic motion be represented by the equation

$$\rho = a \cos pt + k,$$

where a is the amplitude of the vibration, $2\pi/p$ the period, k the distance of the mean point of vibration from the origin of coördinates and t the time. The rotatory motion of a line about the origin may be represented by the equation

$$\theta = qt,$$

where q is the rate of rotation. On eliminating t from these two equations we obtain

$$\rho = a \cos \frac{p}{q}\theta + k, \tag{1}$$

the equation of the cyclic-harmonic curve expressed in polar coördinates.

1.3. The foregoing derivation of equation (1) suggests the following convenient method of constructing by points any cyclic-harmonic curve whose equation is given.

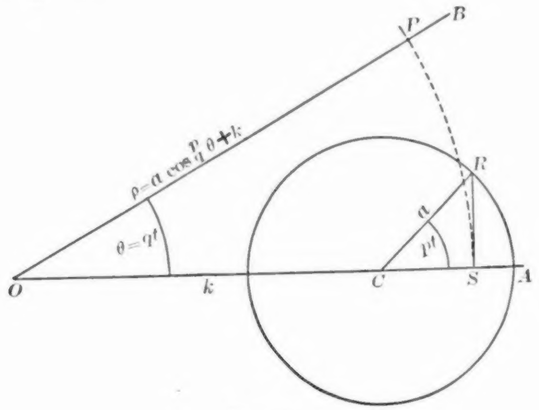


Fig. 1.

Let OA (Fig. 1) represent the initial line, O the pole. From O on OA lay off OC equal to k and with C as a center and a radius equal to a describe the circle of reference of the simple-harmonic motion. Select any convenient unit of angular measure and construct angles ACR and AOB equal to pt and qt units respectively, t being any arbitrarily chosen integer. From R draw RS perpendicular to OA and from O as a center and OS as a radius describe the arc SP cutting OB at P. Then P is a point on the cyclic-harmonic curve $\rho = a \cos (p/q)\theta + k$, for

$$\rho = OP = OS = OC + CS = k + a \cos pt = a \cos \frac{p}{q} \theta + k,$$

since $\theta = qt$.

1.4. By choosing the angular unit sufficiently small, and taking in turn $t = 0, 1, 2, 3$, etc., as many points may be constructed as desired and at intervals small at will. Figures 2, 3, 4, 5 show the method applied to the construction of the cyclic-harmonics $\rho = a \cos 2\theta + k$, for the values $k = 3a$, $k = a$, $k = a/3$, $k = 0$, respectively. Corresponding points on the circle of reference and the cyclic-harmonic are numbered alike.

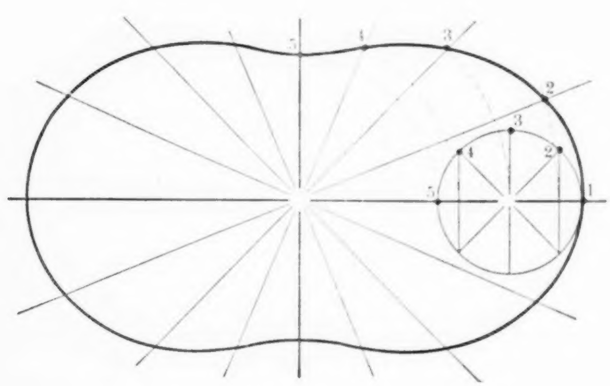


FIG. 2.

1.5. An inspection of the preceding figures discloses certain properties which are independent of the particular value of the ratio p/q employed and which are therefore common to all the species of the genus determined by p/q. Figure 2 has an open center and it follows from the mode of construction, as is otherwise obvious from the form of the equation, that the curve is confined between two circles whose radii are $k - a$ and $k + a$ respectively. Figure 3 consists of leaves which meet in cusps at the origin. The axial diameter of these leaves is $k + a$. Figure 4 consists of two sets of leaves with bases meeting at the origin. The axial diameter of one set of leaves is $k - a$, of the other $k + a$. Figure 5 consists of a single whorl of equal leaves whose axial diameter is a.

1.6. These properties serve as a convenient basis for the classification of the cyclic-harmonics of a given genus p/q. We shall call a cyclic-harmonic curve *curtate* if $k > a$, *cuspitate* if $k = a$, *prolate* if $k < a < 0$, *equi-foliate* if $k = 0$.

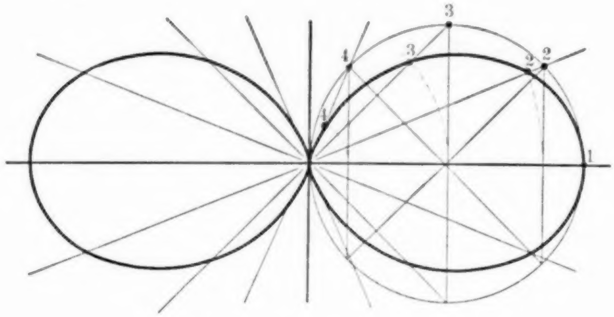


FIG. 3.

1.7. We have assumed the phase of the harmonic motion equal to zero. This restriction may be removed by writing the equation of the harmonic motion in the form

$$\rho = a \cos (pt - \epsilon) + k,$$

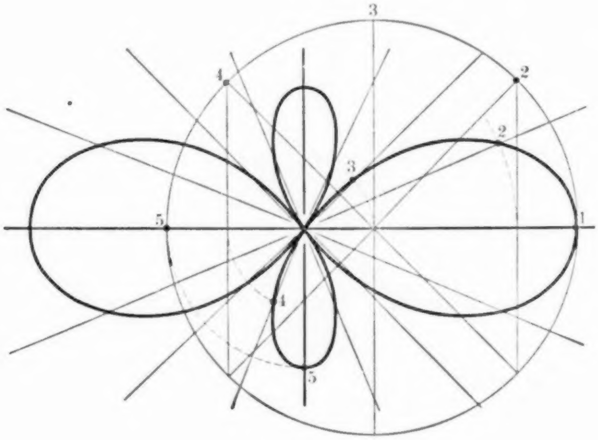


FIG. 4.

where ϵ/p is the phase of the vibration; the equation of the resultant motion then becomes

$$\rho = a \cos \left(\frac{p}{q}\, \theta - \epsilon \right) + k. \tag{2}$$

Equation (2) is only apparently more general than equation (1) for the

former goes over into the latter if we put $\theta = \theta' + \epsilon q/p$, that is, if we turn the initial line through an angle $\epsilon q/p$. There is, therefore, no loss in generality in the curves to be considered if we assume the phase of the vibration equal to zero.

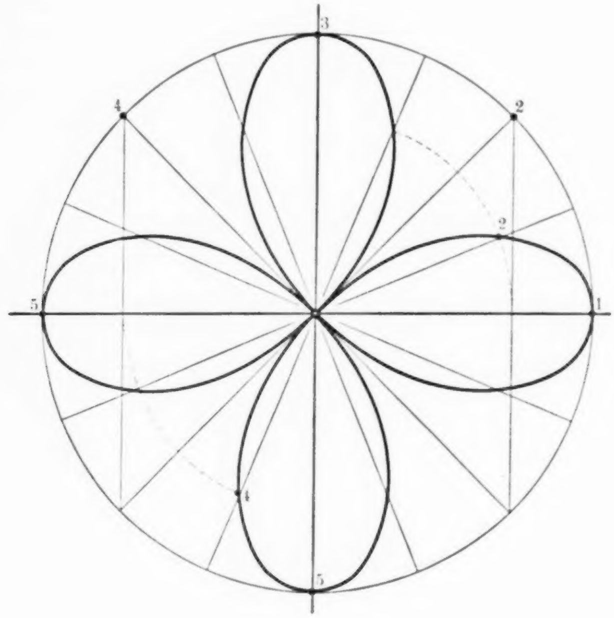


FIG. 5.

1.8. Furthermore, in studying the properties of cyclic-harmonic curves, we may assume a and k both positive. For if a is negative, $a = -a'$,

(3) $\rho = a \cos \frac{p}{q}\theta + k = -a' \cos \frac{p}{q}\theta + k = a' \cos\left(\frac{p}{q}\theta + \pi\right) + k;$

if k is negative, $k = -k'$, then

(4) $\rho = a \cos \frac{p}{q}\theta + k = a \cos \frac{p}{q}\theta - k'$

$$= -\left[a \cos\left(\frac{p}{q}\theta + \pi\right) + k'\right];$$

and if a and k are both negative, $a = -a'$, $k = -k'$, then

(5) $$\rho = a \cos \frac{p}{q}\theta + k = -\left(a' \cos \frac{p}{q}\theta + k'\right).$$

Now (3) is of the form (2), and (4) and (5) differ from (2) and (1) respectively only in the sign of ρ, that is, the curves represented by (3),

(4), and (5) differ from those represented by (1) in position only. In studying the properties of the curves represented by the equation

$$(6) \qquad \rho = a \cos\left(\frac{p}{q}\theta + \epsilon\right) + k, \qquad a, k, \epsilon \text{ positive or negative,}$$

we may therefore, without loss in generality, assume a and k positive and ϵ equal to zero.

1.9. Cyclic-harmonic curves are algebraic or transcendental according as p/q is rational or irrational. For since the curve lies entirely within the circle of radius $a + k$, it will cut a straight line in a finite or infinite number of points according as it does or does not return into itself, that is, according as it is, or is not, possible to satisfy the equation

$$\frac{\rho - k}{a} = \cos\frac{p}{q}\theta = \cos\frac{p}{q}(\theta + 2n\pi), \qquad n \text{ integral.}$$

This equation is satisfied only provided

$$\frac{p}{q}(\theta + 2n\pi) = \frac{p}{q}\theta + 2m\pi, \qquad m \text{ integral,}$$

that is, provided

$$\frac{p}{q} = \frac{m}{n}, \qquad \text{a rational fraction.}$$

We shall throughout the remainder of this paper impose the restriction that p/q be rational and, unless otherwise stated, shall use the term cyclic-harmonic subject to this restriction.

1.10. Cyclic-harmonic curves, as is apparent from their equation, embrace a considerable number of well-known curves. Notable among these are the cardioids and Pascal's limaçon,* Freeth's nephroid,† Münger's double egg curve,‡ and the roses or foliate curves.§ But the simple mode of generation, which gives rise to all of these curves and an infinite number of others, seems to have escaped the observation of previous investigators, there appears not even a record of a common class name or of any attempt at classification. So likewise the many beautiful properties common to all of these curves appear never to have been brought to light, for the reason, no doubt, that in the special cases, which have been carefully studied, these properties are so veiled as to elude detection. It is only from the larger point of view and for the larger values of p and q that these general properties appear in their real significance.

* Roberval, Observations sur la composition des mouvements, Mem. de l'Acad. Royal des Sci., VI, Paris, 1730.

† Proc. Lond. Math. Soc., vol. 10 (1879).

‡ Die eiförmigen Kurven; Dissertation, Bern, 1894.

§ Grandi, Flores geometrici, etc., Florence, 1728. Auth. Dissertation, Marburg, 1866. Hyde, Foliate Curves, The Analyst, II, 1875. Himstedt, Progr. Löbau, 1888.

2. Cyclic-harmonics in Cartesian Coördinates.

2.1. We assume that p/q is positive, rational, and reduced to its lowest terms, so that p is relatively prime to q. Furthermore, unless the contrary is explicitly stated, we shall assume that $k \neq 0$.* The cyclic-harmonic, $\rho = a \cos (p/q)\theta + k$, may then be rationally expressed in terms of Cartesian coördinates, x, y, as follows:

Consider the identity

$$(\cos \theta + i \sin \theta)^p = \cos p\theta + i \sin p\theta = \left(\cos \frac{p}{q}\theta + i \sin \frac{p}{q}\theta\right)^q.$$

If we expand the first member of this identity in terms of powers of $\cos \theta$ and $\sin \theta$ and the last member in terms of powers of $\cos (p/q)\theta$ and $\sin (p/q)\theta$ and then equate the real parts of the two expansions we obtain the new identity

$$\sum_{n=0}^{i}[(-1)^n C_{2n}^p \cos^{p-2n} \theta \sin^{2n} \theta] = \sum_{m=0}^{j}\left[(-1)^m C_{2m}^q \cos^{q-2m}\left(\frac{p}{q}\theta\right)\sin^{2m}\left(\frac{p}{q}\theta\right)\right],$$

where i and j represent the integral parts of $p/2$ and $q/2$ respectively.

Now $\cos \theta = x/\rho$, $\sin \theta = y/\rho$, and from the equation of the cyclic-harmonic curve we have

$$\cos \frac{p}{q}\theta = (\rho - k)/a, \qquad \sin \frac{p}{q}\theta = [a^2 - (\rho - k)^2]^{1/2}/a.$$

On substituting these values in the foregoing identity and multiplying through by the factor $a^q\rho^p$ to avoid fractions, we obtain

$$(7) \qquad a^q\sum_{n=0}^{i}[(-1)^n C_{2n}^p x^{p-2n} y^{2n}] = \rho^p \sum_{m=0}^{j}[(-1)^m C_{2m}^q (\rho - k)^{q-2m} \{a^2 - (\rho - k)^2\}^m],$$

a rational equation between x, y, and ρ.

2.2. The left member of equation (7) is a polynomial, homogeneous of degree p, in x and y. The right member is a polynomial of degree $p + q$ in ρ, consisting of ρ^p multiplied into a polynomial of degree q in $(\rho - k)$. The coefficients of the successive terms $(\rho - k)^q$, $(\rho - k)^{q-2}$, $(\rho - k)^{q-4}$, $\cdots$, $(\rho - k)^{q-2m}$, are

$$B_q, \ -a^2B_{q-2}, \ +a^4B_{q-4}, \ \cdots, \ (-1)^m a^{2m} B_{q-2m},$$

where

$$B_q = 1 + C_2^q + C_4^q + C_6^q + \cdots,$$

* It will be shown that $k = 0$ gives rise to degenerate forms for which many of the general theorems here deduced break down. Nor is it necessary to consider these special cases at length since their properties have been repeatedly investigated. See Loria, Spezielle algebraische und transscendente ebene Kurven, Leipzig (1902), Absch. 5, Kap. 8.

$$\begin{aligned}
B_{q-2} &= C_2^q + C_1^2C_4^q + C_2^3C_6^q + C_3^4C_8^q + \cdots, \\
B_{q-4} &= C_4^q + C_1^3C_6^q + C_2^4C_8^q + C_3^5C_{10}^q + \cdots, \\
&\quad \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ , \\
B_{q-2m} &= C_{2m}^q + C_1^{m+1}C_{2m+2}^q + C_2^{m+2}C_{2m+4}^q + C_3^{m+3}C_{2m+6}^q + \cdots.
\end{aligned}$$

It is now easy to write out the coefficients of the various powers of ρ in the expanded form of the right member of equation (7). Denoting the coefficient of ρ^{p+r} by C_r we have

$$\begin{aligned}
C_q &= B_q, \\
C_{q-1} &= -\,kC_1^qB_q, \\
C_{q-2} &= k^2C_2^qB_q - a^2B_{q-2}, \\
C_{q-3} &= -\,k^3C_3^qB_q + ka^2C_1^{q-2}B_{q-2}, \\
&\qquad \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ , \\
C_{q-r} &= k^rC_r^qB_q - k^{r-2}a^2C_{r-2}^{q-2}B_{q-2} + k^{r-4}a^4C_{r-4}^{q-4}B_{q-4} - \cdots \\
&\qquad + (-1)^{r/2}a^rB_{q-r}, \\
&= -\,[k^rC_r^qB_q - k^{r-2}a^2C_{r-2}^{q-2}B_{q-2} + k^{r-4}a^4C_{r-4}^{q-4}B_{q-4} - \cdots \\
&\qquad + (-1)^{(r-1)/2}ka^{r-1}C_1^{q-r+1}B_{q-r+1}]
\end{aligned}$$

according as r is even or odd. We note in particular that

$$(8)\qquad C_q = [(1+1)^q + (1-1)^q]/2 = 2^{q-1}, \qquad C_{q-1} = -\,kq\cdot 2^{q-1},$$

$$\begin{aligned}
& C_0 = \pm\,[k^qB_q - k^{q-2}a^2B_{q-2} + k^{q-4}a^4B_{q-4} - \cdots \\
(9)\qquad & \qquad = \frac{\pm\,[(k + \sqrt{k^2 - a^2})^q + (k - \sqrt{k^2 - a^2})^q]}{2}.
\end{aligned}$$

2.3. It is obvious that the coefficients of the terms in ρ of the right member of equation (7) are independent of p, that is, these coefficients are invariant for all cyclic-harmonics having the same value of q. Similarly, the coefficients of the terms in x and y of the left member are, barring the factor a^q, independent of q, hence these coefficients are invariant for all values of q. This property greatly facilitates the computation of the Cartesian equations of the various genera of cyclic-harmonics. It should further be observed that the computation of the B's which enter C's may be expedited by superimposing one on the other the two arrays

$$\begin{array}{cccccc}
C_0^q & C_2^q & C_4^q & C_6^q & C_8^q & \cdot \\
C_2^q & C_4^q & C_6^q & C_8^q & C_{10}^q & \cdot \\
C_4^q & C_6^q & C_8^q & C_{10}^q & C_{12}^q & \cdot \\
C_6^q & C_8^q & C_{10}^q & C_{12}^q & C_{14}^q & \cdot \\
C_8^q & C_{10}^q & C_{12}^q & C_{14}^q & C_{16}^q & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot &
\end{array}
\qquad
\begin{array}{cccccc}
C_0^0 & C_1^1 & C_2^2 & C_3^3 & C_4^4 & \cdot \\
C_0^1 & C_1^2 & C_2^3 & C_3^4 & C_4^5 & \cdot \\
C_0^2 & C_1^3 & C_2^4 & C_3^5 & C_4^6 & \cdot \\
C_0^3 & C_1^4 & C_2^5 & C_3^6 & C_4^7 & \cdot \\
C_0^4 & C_1^5 & C_2^6 & C_3^7 & C_4^8 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{array}$$

Any required value B_{q-2m} may then be obtained by adding the products of the superimposed terms in the $(n+1)$th row or column.

2.4. The following table contains the coefficients B_{q-2m} for all values of q from 1 to 10 inclusive.

B_{q-2m}

$m \backslash q$	1	2	3	4	5	6	7	8	9	10
0	1	2	4	8	16	32	64	128	256	512
1		1	3	8	20	48	112	256	576	1280
2				1	5	18	56	160	432	1120
3						1	7	32	120	400
4								1	9	50
5										1

2.5. The foregoing table forms the basis for the computation of the coefficients C_{q-r} which are tabulated below.

C_{q-r}

$r \backslash q$	1	2	3	4	5	6	7	8	9	10	
0	1	2	4	8	16	32	64	128	256	512	
1	1	4	12	32	80	192	448	1024	2304	5120	$\times -k$
2		2	12	48	160	480	1344	3584	9216	23040	$\times k^2$
		1	3	8	20	48	112	256	576	1280	$\times -a^2$
3			4	32	160	640	2240	7168	21504	61440	$\times -k^3$
			3	16	60	192	560	1536	4032	10240	$\times +ka^2$
4				8	80	480	2240	8960	32256	107520	$\times k^4$
				8	60	288	1120	3840	12096	35840	$\times -k^2a^2$
				1	5	18	56	160	432	1120	$\times +a^4$
5					16	192	1344	7168	32256	129024	$\times -k^5$
					20	192	1120	5120	20160	71680	$\times k^3a^2$
					5	36	168	640	2160	6720	$\times -ka^4$
6						32	448	3584	21504	107520	$\times k^6$
						48	560	3840	20160	89600	$\times -k^4a^2$
						18	168	960	4320	16800	$\times +k^2a^4$
						1	7	32	120	400	$\times -a^6$
7							64	1024	9216	61440	$\times -k^7$
							112	1536	12096	71680	$\times +k^5a^2$
							56	640	4320	22400	$\times -k^3a^4$
							7	64	360	1600	$\times +ka^6$
8								128	2304	23040	$\times k^8$
								256	4032	35840	$\times -k^6a^2$
								160	2160	16800	$\times +k^4a^4$
								32	360	2400	$\times -k^2a^6$
								1	9	50	$\times +a^8$
9									256	5120	$\times -k^9$
									576	10240	$\times +k^7a^2$
									432	6720	$\times -k^5a^4$
									120	1600	$\times +k^3a^6$
									9	100	$\times -ka^8$
10										512	$\times k^{10}$
										1280	$\times -k^8a^2$
										1120	$\times +k^6a^4$
										400	$\times -k^4a^6$
										50	$\times +k^2a^8$
										1	$\times -a^{10}$

2.6. The foregoing table of coefficients enables us to write the Cartesian equation of each of the 63 genera of cyclic-harmonic curves, p, $q \equiv 10$, in decreasing powers of ρ. The degree of the equation in ρ is $p + q$ for by (8) the coefficient ρ^{p+q} is different from zero. Now $\rho^2 = x^2 + y^2$, hence the equation is not rational in x and y unless only even powers of ρ occur, but it appears from (8) that for $k \neq 0$ at least one odd power of ρ is present in the equation. One quadrature is therefore necessary and sufficient to rationalize the equation.

2.7. The process of rationalization is exceedingly laborious for all but the lower values of q. It required several days intense work on the part of the writer to compute the coefficients in the Cartesian equation whose equivalent polar form is $\rho = a \cos (\theta/10) + k$. Written at length this equation is:

$$\begin{aligned}
&262{,}144(x^2+y^2)^{11}-(2{,}621{,}440k^2+1{,}310{,}720a^2)(x^2+y^2)^{10}+(11{,}796{,}480k^4\\
&+9{,}175{,}040k^2a^2+2{,}785{,}280a^4)(x^2+y^2)^9-(31{,}457{,}280k^6+26{,}214{,}400k^4a^2\\
&+13{,}107{,}200k^2a^4+3{,}276{,}800a^6)(x^2+y^2)^8+(55{,}050{,}240k^8+36{,}700{,}160k^6a^2\\
&+22{,}937{,}600k^4a^4+9{,}830{,}400k^2a^6+2{,}329{,}600a^8)(x^2+y^2)^7-(66{,}060{,}288k^{10}\\
&+18{,}350{,}080k^8a^2+18{,}350{,}080k^6a^4+9{,}830{,}400k^4a^6+4{,}147{,}200k^2a^8\\
&+1{,}025{,}024a^{10})(x^2+y^2)^6+(55{,}050{,}240k^{12}-18{,}350{,}080k^{10}a^2+11{,}468{,}800k^8a^4\\
&+3{,}276{,}800k^6a^6+2{,}176{,}000k^4a^8+977{,}920k^2a^{10}+274{,}560a^{12})(x^2+y^2)^5\\
&-(31{,}457{,}280k^{14}-36{,}700{,}160k^{12}a^2+18{,}350{,}080k^{10}a^4-3{,}276{,}800k^8a^6\\
&+716{,}800k^6a^8+215{,}040k^4a^{10}+120{,}320k^2a^{12}+42{,}240a^{14})(x^2+y^2)^4\\
&+(11{,}796{,}480k^{16}-26{,}214{,}400k^{14}a^2+22{,}937{,}600k^{12}a^4-9{,}830{,}400k^{10}a^6\\
&+2{,}176{,}000k^8a^8-215{,}040k^6a^{10}+19{,}200k^4a^{12}+6{,}400k^2a^{14}+3{,}300a^{16})(x^2+y^2)^3\\
&-(2{,}621{,}440k^{18}-9{,}175{,}040k^{16}a^2+13{,}107{,}200k^{14}a^4-9{,}830{,}400k^{12}a^6\\
&+4{,}147{,}200k^{10}a^8-977{,}920k^8a^{10}+120{,}320k^6a^{12}-6{,}400k^4a^{14}+200k^2a^{16}\\
&+100a^{18})(x^2+y^2)^2+(262{,}144k^{20}-1{,}310{,}720k^{18}a^2+2{,}785{,}280k^{16}a^4\\
&-3{,}276{,}800k^{14}a^6+2{,}329{,}600k^{12}a^8-1{,}025{,}024k^{10}a^{10}+274{,}560k^8a^{12}\\
&-42{,}240k^6a^{14}+3{,}300k^4a^{16}-100k^2a^{18}+a^{20})(x^2+y^2)-10{,}240ka^{10}x(x^2+y^2)^5\\
&-(122{,}880k^2-20{,}480a^2)ka^{10}x(x^2+y^2)^4-(258{,}048k^4-143{,}360k^2a^2\\
&+13{,}440a^4)ka^{10}x(x^2+y^2)^3-(122{,}880k^6-143{,}360k^4a^2+44{,}800k^2a^4\\
&-3{,}200a^6)ka^{10}x(x^2+y^2)^2-(10{,}240k^8-20{,}480k^6a^2+13{,}440k^4a^4-3{,}200k^2a^6\\
&+200a^8)ka^{10}x(x^2+y^2)-a^{20}x^2=0.
\end{aligned}$$

2.8. We have seen that every cyclic-harmonic curve, $k \neq 0$, leads to an algebraic equation of degree $p + q$ in the ρ's which requires one quadrature in order to make the equation rational in x and y, therefore,

Every cyclic-harmonic curve, $\rho = a \cos (p/q)\theta + k$, $k \neq 0$, is an algebraic curve of order $2(p + q)$.

2.9. If $k = 0$ the second member of (7) reduces to

$$\rho^p\Sigma[(-1)^m C_{2m}\rho^{q-2m}(a^2 - \rho^2)^m],$$

whose degree in ρ is odd or even according as $p + q$ is odd or even. If $p + q$ is even, that is, if both p and q are odd, the equation is rational in x and y as it stands, hence

Cyclic-harmonic curves for which both p and q are odd, and $k = 0$, are algebraic curves of order $p + q$.

Every cyclic-harmonic curve is, therefore, an algebraic curve of even order.

2.10. By considering the rationalized form of the equation of the general cyclic-harmonic curve we see that when both p and q are odd and $k = 0$ this equation reduces to the product of two equal factors. The cyclic-harmonic curve corresponding to this case consists of two equal and coincident branches each of order $p + q$. We shall call a single branch of such a curve a *degenerate* cyclic-harmonic curve. It is obvious that all results derived for the general case will require modification before they can be applied to the degenerate case. Unlike degenerate forms of many other plane curves, degenerate cyclic-harmonic curves cannot be readily recognized by their form alone.

2.11. The number of genera of cyclic-harmonic curves of a given order is readily determined as follows. Let $2n$ be the given order, then the number in question is evidently the number of integral solutions, p, q, of the equation $p + q = n$ subject to the restriction that p, q and n be relatively prime to each other. It follows that any pair of integers, p and $n - p$, of which p is less than n and relatively prime to n, constitute a solution, for if p is relatively prime to n so also is $n - p$. The number sought is, therefore, the *totient function* $\varphi(n)$, and we have the theorem, *The number of genera of cyclic-harmonic curves, $k \neq 0$, having a given order $2n$ is the totient number $\varphi(n)$.*

The foregoing enumeration does not include degenerate cyclic-harmonics. Of such there are $\varphi(2n)$ having a given order $2n$.

The number of genera of cyclic-harmonic curves, $k \neq 0$, whose order is $2n$ or less is $\sum_{n=1}^{n}\varphi(n)$. Besides these there are $\sum_{n=1}^{n}\varphi(2n)$ degenerate forms of order $2n$ or lower order.

2.12. Another interesting inquiry concerns the number of genera of cyclic-harmonic curves for which neither p nor q exceeds a given number n. Suppose first $p > q$, then for a given value of p there are $\varphi(p)$ admissible values of q and hence of the ratio p/q. Now p may take all values from 1 to n inclusive, hence there are $\sum_{n=1}^{n}\varphi(n)$ genera subject to the restriction $n \geqq p > q$. Evidently there is an equal number subject to the restriction $n \geqq q > p$. Finally there is the case $p = q = 1$. Hence

The number of genera of cyclic-harmonic curves, $k \neq 0$, subject to the condition that neither p nor q shall exceed a given number n is $2\sum_{n=1}^{n}\varphi(n) + 1$.

2.13. Returning to equation (7) we see that the order of the terms of lowest degree in x and y is p, after rationalization $2p$, hence

Every cyclic-harmonic curve, $k \neq 0$, has a multiple point of order $2p$ at the origin.

Every degenerate cyclic-harmonic curve has a multiple point of order p at the origin.

2.14. The terms of highest order in the rationalized form of equation (7) result from the expansion of

$$\rho^{2(p+q)} = (x^2 + y^2)^{p+q} = (x + iy)^{p+q}(x - iy)^{p+q};$$

each of the circular rays $x + iy = 0$, $x - iy = 0$, must therefore intersect the curve in $p + q$ coincident points on the line at infinity, hence

Every cyclic-harmonic curve, $k \neq 0$, has $p + q - 1$ fold contact with the line at infinity at each of the circular points.

The line at infinity is thus seen to be a double tangent to every cyclic-harmonic curve; it is an ordinary or inflectional tangent according as $p + q$ is even or odd; moreover since the order of the curve is $2(p + q)$ it can meet the line at infinity in no points other than the circular points.

2.15. That cyclic-harmonic curves are unipartite is sufficiently obvious from their definition; that they are also unicursal or rational may be shown as follows. Since $\rho = a \cos (p/q)\theta + k$,

$$x = \left(a \cos \frac{p}{q} \theta + k\right) \cos \theta, \qquad y = \left(a \cos \frac{p}{q} \theta + k\right) \sin \theta.$$

Now let $\theta/q = \varphi$, then $\theta = q\varphi$, and we have

$$x = (a \cos p\varphi + k) \cos q\varphi, \qquad y = (a \cos p\varphi + k) \sin q\varphi.$$

Now p and q being integers, $\cos p\varphi$ $\cos q\varphi$ and $\sin q\varphi$, may each be rationally expressed in terms of $\sin \varphi$ and $\cos \varphi$ so that on making the substitution

$$\cos \varphi = (1 - t^2)/(1 + t^2), \qquad \sin \varphi = 2t/(1 + t^2), \qquad t = \tan (\varphi/2),$$

x and y may each be expressed rationally in terms of the single parameter t, hence

All cyclic-harmonic curves are rational or unicursal curves.

MORE THEOREMS ON THE COMPLETE QUADRILATERAL.

By J. W. Clawson.

In a paper on "The Complete Quadrilateral" published in these Annals,* a number of theorems were given. These divide themselves naturally into theorems in connection with (A) the *circumcentric circle*, **C**, determined by the circumcenters of the four triangles of the quadrilateral, (B) the *mid-diagonal line*, m, determined by the middle points of lines joining opposite vertices, (C) the *orthocentric line*, o, determined by the four orthocenters, and the *pedal line*, p, which are both perpendicular to m, (D) the *incentric lines*, which are connected with the bisectors of angles of the quadrilateral. These four divisions of the subject are somewhat loosely connected, in the paper referred to, by the facts that the *focal point*, F, at which the four circumcenters meet, (A) lies on **C**, (B) is the focus of the most important of the conics whose centers are on m, (C) is simply related to p, and (D) is the intersection of the incentric lines.

In this note some further connective theorems are added. In 1 and 2, (A), (B) and (C) are more closely linked, in 3, (A) and (D) are bound together, and in 4, relations are given connecting (A), (B) and (D).

The notation of my former paper is preserved, and most of the references are to it. The contents of this note are original, except where otherwise stated.

1. (1) The mid-diagonal line, m, of the quadrilateral bisects the line joining the center, C, of the circumcentric circle and the mean center, H, of the four orthocenters of the triangles of the quadrilateral.

I have discovered two proofs of this theorem, both too long for insertion in full. The first is analytical. Taking the focal point for origin, and taking the equation of the line l_1 to be $px + q_1y = p^2 + q_1^2$, the mid-diagonal line is found, after considerable reduction, to have for its equation $y = \frac{1}{2}\Sigma q_1$, the point C, the center of the circumcentric circle, is

$$\left(\frac{p^4 - p^2\Sigma q_1q_2 + q_1q_2q_3q_4}{4p^3}, \frac{p^2\Sigma q_1 - \Sigma q_1q_2q_3}{4p^2}\right),$$

and the point H, the mean center or center of gravity of equal masses

* Vol. 20, pp. 232–261.

placed at H_1, H_2, H_3, H_4, is

$$\left(2p, \frac{3p^2\Sigma q_1 + \Sigma q_1q_2q_3}{4p^2}\right).$$

Hence the middle point of CH has $\frac{1}{2}\Sigma q_1$ for its ordinate.

The second proof is statical. It can be proved, using trigonometrical methods, that masses $\sin 2A_{23}$ at C_1, $\sin 2A_{13}$ at C_2, $\sin 2A_{12}$ at C_3, and $\sin A_{23} \sin A_{13} \sin A_{12}$ at each of the four points H_1, H_2, H_3, H_4 are equivalent to masses $4 \sin A_{23} \sin A_{13} \sin A_{12}$ at C and at H, and hence to the single mass $8 \sin A_{23} \sin A_{13} \sin A_{12}$ at the middle point of CH.

Again, the masses at C_1 and H_1 may be replaced by certain masses at the vertices of the triangle $A_{23}A_{24}A_{34}$ whose circumcenter and orthocenter are C_1 and H_1. In this way the seven original masses may be replaced by masses at the six vertices; and it can be proved, laboriously, that the masses at opposite vertices are equal; hence that the seven masses are replaceable by three masses at B_1, B_2, B_3, the middle points of the diagonals. But the centroid of these three masses is at a point on the mid-diagonal line. Hence the middle point of CH lies on this line.

(2) U, the center of gravity of equal masses placed at the six vertices of the quadrilateral, is the centroid of the triangle whose vertices are C, H, and the orthic center,* O, of the quadrangle $C_1C_2C_3C_4$.

This is easily established statically. For masses $2m$ at each of the six vertices may be replaced by $3m$ at the centroid of each of the four triangles of the quadrilateral. But, since the centroid of a triangle is one third of the distance from the circumcenter to the orthocenter, these may be replaced by four masses of m each at H_1, H_2, H_3, H_4 and four masses of $2m$ each at C_1, C_2, C_3, C_4. Now the mean center† of the quadrangle $C_1C_2C_3C_4$ bisects‡ the line joining C, its center, to O, its orthic center,—the point where perpendiculars to each side from the middle point of the opposite side concur.* Hence these masses may be replaced by $4m$ at H, and $4m$ at C and $4m$ at O. But the six masses of $2m$ each at the vertices may also be replaced by a mass of $12m$ at U. Hence U is the centroid of equal masses at H, C and O.

Since OU produced bisects CH, and since, by (1), the mid-diagonal line, which contains U,§ bisects CH, it follows that:

(3) The mid-diagonal line of a complete quadrilateral contains the orthic center of the circumcentric quadrangle.

2. Let h_1, h_2, h_3, h_4 be the orthocenters of the triangles $C_2C_3C_4$, $C_3C_4C_1$,

* P. 252 (γ), Annals, l.c.

† P. 251 (α).

‡ P. 252 (δ).

§ P. 238 (9).

$C_4C_1C_2$, $C_1C_2C_3$ respectively. Then C_1h_1, C_2h_2, C_3h_3, C_4h_4 are bisected at O,* and the quadrangle $h_1h_2h_3h_4$ is directly similar to the circumcentric quadrangle $C_1C_2C_3C_4$. Let K be the center of the circle circumscribing $h_1h_2h_3h_4$. Then CK is bisected at O. Also KC_4 is equal and parallel to h_4C.

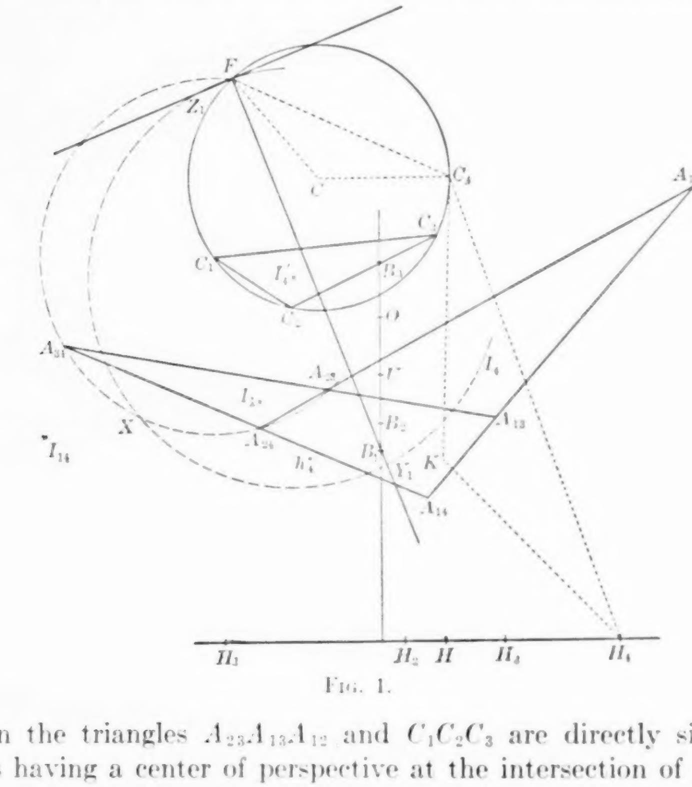


FIG. 1.

Again the triangles $A_{23}A_{13}A_{12}$ and $C_1C_2C_3$ are directly similar, the triangles having a center of perspective at the intersection of $\mathbf{C}_4$ and $\mathbf{C}$† other than F, viz., E_4. The orthocenters of these triangles are H_4, h_4; their circumcenters are C_4, C; their radii are R_4, R. Then $H_4C_4/h_4C = R_4/R$. Hence $H_4C_4/KC_4 = R_4/R$.

Now consider the triangles H_4C_4K and C_4FC. By the last statement, $H_4C_4/KC_4 = C_4F/CF$. But $\measuredangle\, KC_4H_4$ is equal to the angle between h_4C and H_4C_4. But h_4C makes the same angle with C_2C_3 that H_4C_4 makes with $A_{13}A_{12}$, considering the similar figures. Hence $\measuredangle\, KC_4H_4$ is equal to the angle between C_2C_3 and $A_{13}A_{12}$. But C_2C_3 is perpendicular to FA_{14}. Hence $\measuredangle\, KC_4H_4$ is the complement of $\measuredangle\, A_{13}A_{14}F$.

Again $\measuredangle\, CFC_4$ is the complement of $\measuredangle\, C_4C_2F$, i.e., of $\measuredangle\, A_{13}A_{14}F$.

Hence the above-named triangles are similar. But triangle C_4FC is isosceles. Hence KC_4 is equal to KH_4. Thus K lies on the perpendicular bisector of C_4H_4.

* P. 253 (e).

† P. 235 (5).

Similarly K lies on the perpendicular bisectors of C_1H_1, C_2H_2, C_3H_3. This gives a new proof of Hervey's theorem* that

(4) The perpendicular bisectors of the lines joining the circumcenters and orthocenters are concurrent.

It also connects this point with other points of the quadrilateral, since

(5) The line joining this point of concurrence to the center of the circumcentric circle is bisected by the orthic center of the circumcentric quadrangle.

Moreover

(6) HK is parallel to the mid-diagonal line.

3. Starting† from the fact that the sixteen centers of the circles inscribed and escribed to the four triangles of the quadrilateral are four by four concyclic, giving rise to eight new circles, whose centers Y_1, Y_2, Y_3, Y_4 and Z_1, Z_2, Z_3, Z_4 are on the incentric lines,‡ it is easy to see that

(7) The circle on Y_1Z_1 as diameter passes through the points common to the (orthogonal) circles just named whose centers are at Y_1 and Z_1. One of these points is I_4. This circle also passes through F and through the middle points of I_1I_{14}, $I_{21}I_{23}$, I_3I_{34}, as is easily proved. There are sixteen circles of this kind which all pass through F.

But it is more remarkable that

(8) The centers of these sixteen circles are the incenters and excenters of the circumcentric quadrangle $C_1C_2C_3C_4$.

For C_1C_2, C_1C_3 are perpendicular respectively to FA_{34}, FA_{24}. Hence the bisector of $\measuredangle\ C_2C_1C_3$ is perpendicular to the bisector of $\measuredangle\ A_{34}FA_{24}$. Now the circle $\mathbf{C}_1$ is the nine-point circle of the triangle $I_{12}I_{13}I_{14}$. Let $A_{23}I_1$ cut $\mathbf{C}_1$ at X. Then X is the middle point of I_1I_{14}. Also the bisector of $\measuredangle\ A_{34}FA_{24}$ cuts $\mathbf{C}_1$ at X. Hence the bisector of $\measuredangle\ C_2C_1C_3$ is perpendicular to FX. But the circles Y_1FZ_1 and $\mathbf{C}_1$ have X and F in common. Hence the center of Y_1FZ_1 lies on the bisector of $\measuredangle\ C_2C_1C_3$. Call this point I_1'. In this way the theorem is proved.

From the fact that the middle points of Y_1Z_1, Y_1Z_2, Y_1Z_3, Y_1Z_4 lie on a line parallel to $Z_1Z_2Z_3Z_4$, it follows that

(9) The centers of these sixteen circles lie four by four on four lines parallel to one of the incentric lines and also four by four on lines parallel to the other incentric line.

* P. 244 (25).

† I am indebted to a paper by F. V. Morley in the American Mathematical Monthly for June, 1920 (vol. 27, p. 252), for all the facts contained in this section. Mr. Morley derives the theorems as a special case from a chain of theorems concerning the incenters of n directed lines. It seems worth while to state the theorems in different order and language and to derive them by pure geometry from simpler rather than from more complex theorems.

‡ P. 245 (27), p. 246 (27a).

Moreover

(10) The incentric lines of the quadrilateral are parallel to the bisectors of the angles between pairs of opposite sides of the circumcentric quadrangle.*

4. If the figure that we are considering is inverted with respect to the center F, I have shown elsewhere† that a new figure results which is inversely similar to the old one, one of the incentric lines being an axis of similitude; and that the circumcentric circle and orthocentric line of the old figure invert into the orthocentric line and circumcentric circle of the new one, while the incentric lines invert into themselves. It follows at once from these facts that

(11) FC and the mid-diagonal lines are equally inclined to the incentric lines of the quadrilateral.

It further appears that

(12) If the incentric lines cut the circumcentric circle at F, J and F, J', respectively, the diameter JJ' is parallel to the mid-diagonal line.

URSINUS COLLEGE.

* P. 256 (ω).

† Amer. Math. Monthly, vol. 24, (1917), p. 71.

A THEOREM ON CROSS-RATIOS IN THE GEOMETRY OF INVERSION.

By J. L. Walsh.

It is the purpose of this paper to present a solution of the following problem. Let C_1, C_2, C_3, C_4 be four fixed distinct non-null circles in the plane. Let z_1, z_2, z_3, z_4 be the inverses of a variable point z regarding these circles respectively. What are the geometrical characteristics of the configuration of these four circles if the cross-ratio (z_1, z_2, z_3, z_4) is a constant independent of the position of the point z? A complete answer to this question is given in Theorem III below.

As is usual in the geometry of inversion, we adjoin to the finite plane a single point at infinity, and we use the term *circle* to include straight lines as well as circles in the ordinary sense of the word. In proving Theorem III we shall give several preliminary theorems.

Theorem I. *Let α_1, α_2, α_3, α_4 be any four fixed non-concyclic points of the plane. Denote by C_i $(i = 1, 2, 3, 4)$ the circle passing through the three of these points obtained by omitting α_i; and denote by z_i the inverse of a variable point z with respect to C_i. Then the cross-ratio (z_1, z_2, z_3, z_4) does not depend on the position of the point z but is constantly equal to $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.**

By means of a linear transformation, transform α_1 to infinity, α_4 to the point 1, and the inverse of α_1 with respect to C_1 to the origin. This transformation is always possible. The circle C_1 is the unit circle whose center is the origin, and z_2 and z_3 are points on C_1 distinct from each other and from the point 1. The circle C_2 is the line through 1 and α_3, C_3 is the line through 1 and α_2, C_4 is the line through α_2 and α_3.

The function z_i $(i = 1, 2, 3, 4)$ is a linear function of $\bar{z}$, the conjugate imaginary of z. For we may consider z_i obtained from z by reflection in the axis of reals (which brings us to $\bar{z}$) followed by reflection in the circle C_i. Successive reflection in two circles is always a linear transformation.

Let us compute z_i in terms of z. Of course we have $z_1 = 1/\bar{z}$. The other functions z_i are integral functions of $\bar{z}$ since $\bar{z} = \infty$ corresponds to $z_i = \infty$. Moreover $z = z_2$ when $z = 1$ or α_3, $z = z_3$ when $z = 1$ or α_2, $z = z_4$ when $z = \alpha_2$ or α_3. Since $\alpha_2\bar{\alpha}_2 = 1$, $\alpha_3\bar{\alpha}_3 = 1$, we have

$$\begin{cases} z_2 = -\alpha_3\bar{z} + 1 + \alpha_3, \\ z_3 = -\alpha_2\bar{z} + 1 + \alpha_2, \\ z_4 = -\alpha_2\alpha_3\bar{z} + \alpha_2 + \alpha_3. \end{cases}$$

* Strictly speaking, there is an exception if $z = \alpha_i$, for the cross-ratio (z_1, z_2, z_3, z_4) is then not defined. The theorem is true in the sense that whenever the cross-ratio (z_1, z_2, z_3, z_4) is defined, it has the value $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$. A similar remark applies to Theorems II and III below.

The cross-ratio which we are considering is

$$(z_1, z_2, z_3, z_4) \equiv \frac{(z_1 - z_2)(z_3 - z_4)}{(z_2 - z_3)(z_4 - z_1)},$$

which reduces to

$$\frac{\alpha_3 - 1}{\alpha_3 - \alpha_2} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4).$$

This completes the proof.*

It will be noted that the constant cross-ratio referred to in Theorem I is never real. A theorem analogous to Theorem I but referring instead to real cross-ratios is

THEOREM II. *If we denote by z_1, z_2, z_3, z_4 the respective inverses of a variable point z regarding four fixed non-null coaxal circles C_1, C_2, C_3, C_4, then the cross-ratio (z_1, z_2, z_3, z_4) is real and independent of the position of z.*

The four given circles may have two common points, they may all be tangent at a single point, or they may have no common point. We consider these cases in order.

If all four circles have two common points, we transform one of these points to infinity and the other to the origin. The circles C_1, C_2, C_3, C_4 are transformed into straight lines through the origin; we denote by θ_k (where $k = 1, 2, 3, 4$) any angle which the line C_k makes with the axis of reals. The inverse of the point $z = re^{i\varphi}$ with respect to the circle C_k is then $re^{i(2\theta_k - \varphi)}$, and the cross-ratio with which we are concerned is therefore

$$\frac{[re^{i(2\theta_1-\varphi)} - re^{i(2\theta_2-\varphi)}][re^{i(2\theta_3-\varphi)} - re^{i(2\theta_4-\varphi)}]}{[re^{i(2\theta_2-\varphi)} - re^{i(2\theta_3-\varphi)}][re^{i(2\theta_4-\varphi)} - re^{i(2\theta_1-\varphi)}]},$$

which reduces to

$$\frac{[e^{2i\theta_1} - e^{2i\theta_2}][e^{2i\theta_3} - e^{2i\theta_4}]}{[e^{2i\theta_2} - e^{2i\theta_3}][e^{2i\theta_4} - e^{2i\theta_1}]},$$

a number independent of z. In the form in which it is written, this number represents the cross-ratio of the four inverses of $z = 1$. These all lie on the unit circle whose center is the origin and hence their cross-ratio is real.

---

* We indicate briefly another proof of Theorem I. Suppose the configuration transformed as previously indicated. The function $(z_1 - z_2)(z_3 - z_4)/(z_2 - z_3)(z_4 - z_1)$ is a rational function of z. It can become infinite only when $z_2 = z_3$ or $z_4 = z_1$, that is, when $z = \alpha_1, \alpha_2, \alpha_3$, or α_4. For definiteness consider the point α_3, and allow z to lie on the circle C_1 and to approach the point $z = \alpha_3$. The quotient $(z_3 - z_4)/(z_3 - z_2)$ approaches unity, the quotients $(z_1 - z_2)/(z_1 - \alpha_3)$, $(z_1 - \alpha_3)/(z_1 - z_4)$ approach finite limits and hence the rational function does not become infinite at $z = \alpha_3$. Similarly it can be proved not to become infinite at any of the points $z = \alpha_1$, α_2, or α_4, and hence is a constant. It remains to evaluate the constant.

Let $z = 0$, so that $z_1 = \infty$. The points $z_2 = 1 + \alpha_3$, $z_3 = 1 + \alpha_2$, $z_4 = \alpha_2 + \alpha_3$ are the vertices of a triangle congruent to the triangle whose vertices are α_2, α_3, $\alpha_4 = 1$. Then the cross-ratio $(z_1, z_2, z_3, z_4) = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ for $z = 0$ and hence for all values of z.

If the four original circles are all tangent at a single point, we transform that point to infinity and the circles into lines parallel to the axis of imaginaries. The circles C_1, C_2, C_3, C_4 will be lines $x = a_1, a_2, a_3, a_4$ respectively. The inverse of a point $z = x + iy$ with regard to the circle C_k is the point $(2a_k - x) + iy$. The cross-ratio of the four inverses is

$$\frac{(a_1 - a_2)(a_3 - a_4)}{(a_2 - a_3)(a_4 - a_1)},$$

which is not only real and independent of the position of z but is also the cross-ratio of the points in which the lines are cut by any transversal.

If the four original circles have no point in common, there are two null circles of the coaxal family. Transform one of these to infinity and the other to the origin, so that C_1, C_2, C_3, C_4 become circles whose common center is the origin; we denote their respective radii by r_1, r_2, r_3, r_4. The inverse of $z = re^{i\varphi}$ with regard to C_k is $r_k^2 e^{i\varphi}/r$, and the cross-ratio of the four inverses reduces to

$$\frac{(r_1^2 - r_2^2)(r_3^2 - r_4^2)}{(r_2^2 - r_3^2)(r_4^2 - r_1^2)},$$

which is real and independent of z. The proof of Theorem II is thus complete.

Suppose now we have four distinct fixed non-null circles C_1, C_2, C_3, C_4, and that the cross-ratio (z_1, z_2, z_3, z_4) of the four inverses of a point z is a constant independent of z. We shall prove that we have a situation such as appears either in Theorem I or Theorem II.

The cross-ratio can have none of the degenerate values $0, 1, \infty$. Suppose for definiteness that it has the constant value zero. Then either $z_1 = z_2$ or $z_3 = z_4$ for an infinite number of values of z. From the reasoning previously used, z_1 is a linear function of z_2 and z_3 is a linear function of z_4. Hence we must have $z_1 \equiv z_2$ or $z_3 \equiv z_4$, which means that C_1 coincides with C_2 or C_3 coincides with C_4; either of these suppositions is contrary to our hypothesis.

If any two of the circles C_1, C_2, C_3, C_4, say for definiteness C_1 and C_2, have a point α in common, a third circle of the set must pass through α. For we may choose $z = \alpha$, so that $z_1 = z_2 = \alpha$. If the cross-ratio is not to have the value zero we must have either $z_2 = z_3 = \alpha$ or $z_4 = z_1 = \alpha$; a point coincides with its inverse only when it is on the circle of inversion, so C_3 or C_4 must pass through α.

If any two of the original four circles, for definiteness C_1 and C_2, have no point in common, a third circle of the set is coaxal with them. For there exist two points α and β mutually inverse regarding both C_1 and C_2. Let $z = \alpha$ and we have $z_1 = z_2 = \beta$. Hence we must have $z_2 = z_3 = \beta$

or $z_4 = z_1 = \beta$; that is, the inverse of α with respect to C_3 or C_4 is β. Then C_3 or C_4 is coaxal with C_1 and C_2.

If two of the circles C_1, C_2, C_3, C_4 have no point in common, all four circles are coaxal. For we have just shown three—for definiteness C_1, C_2, C_3—to be coaxal; no two of these three circles can have a point in common. Then C_4 can have no point in common with any of the circles C_1, C_2, C_3. If it has no point in common with C_1, the circles C_1, C_2, C_4 or C_1, C_3, C_4 must be coaxal. Hence all four circles are coaxal.

If two, say C_1 and C_2, of the original four circles are tangent at a point α, all four circles are mutually tangent at α. If C_3 is not tangent to C_1 and does not pass through α, it must cut C_1 in two points distinct from α. Then C_4 must pass through α and through these two points and hence coincide with C_1. Therefore each of the circles C_3 and C_4 must either pass through α or be tangent to both C_1 and C_2 in points distinct from α. Both C_3 and C_4 cannot pass through α unless both are tangent at α to C_1 and C_2; both C_3 and C_4 cannot be tangent to C_1 and C_2 in points distinct from α. Suppose for definiteness that C_4 does not pass through α but is tangent to C_1 and C_2 respectively in points β and γ distinct from α. Then C_3 must pass through α, β, and γ. Transform α, β, γ to ∞, $+1$, -1, respectively. Then C_3 becomes the axis of reals, C_4 becomes the unit circle whose center is the origin, and C_1 and C_2 become the lines tangent to C_4 at $+1$ and -1 respectively. These four circles do not satisfy the hypothesis we have made. For when z is on C_3, all the points z_1, z_2, z_3, z_4 are also on C_3 and hence their cross-ratio is real. On the other hand, if z is not on C_3 but is on C_1, z_4 is interior to the triangle formed by z_1, z_2, z_3, the four points are not concyclic, their cross-ratio is not real and therefore not constant.

If two of the circles C_1, C_2, C_3, C_4, say C_1 and C_2, have two distinct points α and β in common, the four circles either are coaxal or form a configuration such as that described in Theorem I. For either C_3 or C_4, say C_3, must pass through α. If C_3 passes through β as well, C_4 must pass through both α and β and hence the statement is proved. If C_3 passes through α but not through β, it intersects C_1 and C_2 respectively in points γ and δ distinct from each other and from α and β. Then C_4 must pass through β, γ, and δ, so we have the kind of configuration described in Theorem I. This completes the proof of

THEOREM III. *Let C_1, C_2, C_3, C_4 be four distinct fixed non-null circles. Denote by z_1, z_2, z_3, z_4 the inverses of a variable point z with regard to these four circles respectively. A necessary and sufficient condition that the cross-ratio (z_1, z_2, z_3, z_4) be real and independent of the position of z is that C_1, C_2, C_3, C_4 be coaxal. A necessary and sufficient condition that the cross-ratio*

*be non-real and independent of the position of z is that the four circles pass by threes through four distinct points. If we denote by α_i the point through which pass the three circles which do not include C_i, we shall have**

$$(z_1, z_2, z_3, z_4) \equiv (\alpha_1, \alpha_2, \alpha_3, \alpha_4).$$

In Theorem III we have supposed none of the circles C_1, C_2, C_3, C_4 to be a null circle. We shall now consider the possibility that some or all of these may be null circles, but shall suppose that all four circles are distinct. It follows as before that the cross-ratio does not degenerate.

If we choose any four points of the plane, α_1, α_2, α_3, α_4, consider them as null circles, and consider the inverse of a point z with regard to α_i to be the point α_i itself, of course the cross-ratio (z_1, z_2, z_3, z_4) is constantly $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$. Three of the four original circles cannot be null circles unless the fourth is also a null circle. For three of the points z_1, z_2, z_3, z_4 and a constant cross-ratio determine uniquely the fourth of those points, which is therefore fixed independent of z.

Suppose two of the original circles, for definiteness C_1 and C_2, are null circles while the other two are non-null circles. We consider in detail the possibilities that C_3 and C_4 have two points in common, are tangent, or have no point in common. If C_3 and C_4 have two points in common, the proof formerly given shows that C_1 and C_2 must lie at the two intersections of C_3 and C_4. Transform C_1 and C_2 to the origin and to infinity respectively, and denote by θ_3 and θ_4 the respective angles which C_3 and C_4—now straight lines through the origin—make with the axis of reals. If we choose any point $z = r^{i\varphi}$, the corresponding inverses are $z_1 = 0$, $z_2 = \infty$, $z_3 = re^{i(2\theta_3-\varphi)}$, $z_4 = re^{i(2\theta_4-\varphi)}$. The cross-ratio is

$$(z_1, z_2, z_3, z_4) = \frac{\epsilon^{2i\theta_4} - \epsilon^{2i\theta_3}}{\epsilon^{2i\theta_4}},$$

which is independent of the position of the point z.

If C_3 and C_4 are tangent, their point of tangency must be either C_1 or C_2, say for definiteness C_1. Transform C_1 to infinity, C_2 to the origin, and C_3 and C_4 into lines parallel to the axis of imaginaries. When z is real, its four inverses are concyclic and hence their cross-ratio is real. When z is not real, the four inverses are not concyclic and their cross-ratio is not real and hence not constant.

If C_3 and C_4 have no point in common, the proof formerly given shows that C_1, C_2, C_3, C_4 are coaxal. Hence C_1 and C_2 are the null circles of the coaxal family determined by C_3 and C_4. The reader can easily compute the cross-ratio of the four inverses of z and show that it is independent of z.

* Part of the proof of so much of Theorem III as refers to the necessity of the condition was worked out jointly by Professor J. L. Coolidge and myself. Theorem IV was suggested to me by Professor Coolidge.

We give now the results if one and only one of the original circles is a null circle. The proofs are so similar to the foregoing that they are omitted. If two of the original circles have two points in common, a third is coaxal with them and the fourth is a point common to them all. If two of the circles are tangent, a third is coaxal with them and the fourth is the common point of tangency. If two of the circles have no point in common, a third is coaxal with them and the fourth is a null circle of that coaxal family. In all of these cases the cross-ratio (z_1, z_2, z_3, z_4) is a constant independent of the position of the point z.

A theorem closely connected with the first part of Theorem III is the following:

THEOREM IV. *Let there be given four distinct fixed non-null circles in the plane. Denote by z_1, z_2, z_3, z_4 the inverses in these circles respectively of a variable point z of the plane. A necessary and sufficient condition that z_1, z_2, z_3, z_4 be concyclic whatever be the position of z is that the four given circles be coaxal.*

The necessity of the condition is easily proved by methods somewhat similar to those previously used. Denote the given circles by C_1, C_2, C_3, C_4 respectively. Choose z on C_1 but on none of the other circles; then z coincides with z_1. The circle C through z_1, z_2, z_3, z_4 passes through three pairs of distinct points mutually inverse regarding C_2, C_3, C_4, and hence C is orthogonal to C_2, C_3, C_4. We can choose z on C_1 but on none of the circles C_2, C_3, C_4 in an infinite variety of ways and hence we have either an infinity of circles C orthogonal to C_2, C_3, C_4 in which case these three circles are coaxal or we have C coinciding with C_1, so that C_1 is orthogonal to C_2, C_3, C_4. Similarly, we can of course choose C_2, C_3, or C_4 instead of C_1 and prove for example that either C_1, C_3, C_4 are coaxal or C_2 is orthogonal to them all.

If any three of the four circles C_1, C_2, C_3, C_4, say for definiteness C_2, C_3, C_4, are coaxal, all four circles are coaxal. For we know that either C_1, C_3, C_4 are coaxal or C_2 is orthogonal to them all. Since C_2 is coaxal with C_3 and C_4 it is not orthogonal to both C_3 and C_4. Hence C_1, C_3, C_4 are coaxal and therefore all four circles are coaxal.

If no set of three of the four original circles is coaxal, each of those four circles is orthogonal to the other three, which is of course impossible. In fact two of the circles are easily transformed into two perpendicular lines; a third circle must have its center at their intersection; there is evidently no fourth circle orthogonal to all three.

The necessity of the condition of Theorem IV has thus been proved; its sufficiency follows from the reality of the cross-ratio in Theorem II and hence completes the proof. In Theorem IV the four points z_1, z_2,

z_3, z_4 are not only concyclic but are concyclic with z. This follows from inspection of the proof of Theorem II. It immediately suggests

THEOREM V. *Let there be given three distinct fixed non-null circles in the plane. Denote by z_1, z_2, z_3 the inverses in these circles respectively of a variable point z of the plane. A necessary and sufficient condition that z, z_1, z_2, z_3 be concyclic whatever be the position of z is that the three given circles be coaxal.*

First we prove the necessity of the condition. Through any point z not on one of the given circles and through its inverses z_1, z_2, z_3 there passes a circle which is orthogonal to the three given circles. Hence those circles are coaxal.

The sufficiency of the condition follows easily by the method of proof of Theorem II.

HARVARD UNIVERSITY.

THE CONDITION FOR AN ISOTHERMAL FAMILY ON A SURFACE.

By James K. Whittemore.

Consider a real surface and let the rectangular coördinates of its points be given as functions of the two real parameters u, v; suppose the linear element given by

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2.$$

The condition that a family of curves on the surface, $\lambda(u, v) = c$, be isothermal, as generally given, is that $\Delta_2(\lambda)/\Delta_1(\lambda)$ be a function of λ, where $\Delta_1(\lambda)$ and $\Delta_2(\lambda)$ are the first and second differential parameters formed with respect to the linear element of the surface.* This condition is not applicable in the case of frequent occurrence where the family of curves is given not in finite form but by a differential equation. Lie has proved† that if u, v are isothermic parameters in a plane the integral curves of the differential equation $dv/du = \alpha(u, v)$ form an isothermal family when and only when arc tan α is a harmonic function of u and v, and that in this case the equation may be integrated by quadratures. It may be remarked that the theorem is given by Lie as an application of his method of solving a differential equation admitting a known infinitesimal transformation; further, that his proof applies without change to the case of any surface given in terms of isothermic parameters. Lie has also shown‡ that the equation can be integrated by two quadratures if it defines an isothermal family on any surface given with any coördinates, but he has given no method of determining when this is the case.

In this paper we obtain by a simple method, quite different from Lie's, the condition that the differential equation, $dv/du = \alpha$, define an isothermal family on any real surface given with any real parameters u, v, a condition which is a generalization of Lie's condition for isothermic parameters; we prove Lie's theorem that the equation can be integrated by two quadratures when it defines an isothermal family; finally we give the geometrical significance of the angles of the complex integrating factors of the differential equations of the minimal lines of the surface.

Let ω be the angle measured from the positive direction of (v) to the positive direction of (u), where (u) and (v) mean the curves u constant

* See Eisenhart, Differential Geometry (1909), pp. 84, 89, 96.

† Lie-Scheffers, Differentialgleichungen (1891), pp. 156, 157.

‡ L. c., pp. 160–162.

and v constant respectively. Then

$$\sin \omega = \frac{H}{\sqrt{EG}}, \qquad \cos \omega = \frac{F}{\sqrt{EG}},$$

where $\sqrt{EG}$ is positive and H is the positive square root of $EG - F^2$. Let φ be the angle measured in the same direction as ω from the positive direction of (v) to the positive direction on that integral curve C of $dv/du = \alpha$ which passes through the point u, v. The positive directions on (v) and on C are the directions in which the parameter u increases. Considering the infinitesimal triangle whose sides are (v), $(u + du)$, C, we have

$$\frac{\sqrt{G}\,dv}{\sqrt{E}\,du} = \frac{\sqrt{G}}{\sqrt{E}}\alpha = \frac{\sin \varphi}{\sin(\omega - \varphi)} = \frac{\sqrt{EG}}{H \cot \varphi - F},$$

from which

$$(1) \qquad \alpha = \frac{E}{H \cot \varphi - F}, \qquad \tan \varphi = \frac{H\alpha}{E + F\alpha}.$$

The minimal lines of the surface are given by $ds^2 = 0$. Since

$$Eds^2 = (Edu + Fdv)^2 + H^2dv^2$$

the minimal lines are the integral curves of the two equations,

$$(2) \qquad Edu + (F - iH)dv = 0, \qquad Edu + (F + iH)dv = 0.$$

We may assume since E, F, H, u, v are real that integrating factors of equations (2) are respectively $\rho e^{i\theta}$ and $\rho e^{-i\theta}$. Then

$$(3) \qquad \begin{aligned} \rho e^{i\theta}[Edu + (F - iH)dv] &= dx + idy, \\ \rho e^{-i\theta}[Edu + (F + iH)dv] &= dx - idy. \end{aligned}$$

Since these equations give

$$ds^2 = \frac{1}{\rho^2 E}(dx^2 + dy^2)$$

x and y are isothermic parameters of the surface, and are, with a suitable choice of ρ, θ, any pair of isothermic parameters of the surface. Either of equations (3) gives

$$(4) \qquad \begin{aligned} dx &= \rho[(Edu + Fdv)\cos \theta + H \sin \theta dv], \\ dy &= \rho[(Edu + Fdv)\sin \theta - H \cos \theta dv]. \end{aligned}$$

The conditions of integrability for equations (4) are

$$\frac{\partial}{\partial v}(\rho E \cos \theta) = \frac{\partial}{\partial u}(\rho F \cos \theta + \rho H \sin \theta),$$

$$\frac{\partial}{\partial v}(\rho E \sin \theta) = \frac{\partial}{\partial u}(\rho F \sin \theta - \rho H \cos \theta).$$

Expanding and combining the last two equations, we have

$$\frac{\rho_u}{\rho}H + H_u + E\theta_v - F\theta_u = 0, \tag{5}$$

$$\frac{\rho_v}{\rho}EH + H(E_v - F_u - H\theta_u) + F(H_u + E\theta_v - F\theta_u) = 0,$$

where subscripts denote partial differentiation. The condition of integrability of (5), considered as equations in ρ, is

$$\frac{\partial}{\partial v}\left[\frac{H_u + E\theta_v - F\theta_u}{H}\right]$$
$$= \frac{\partial}{\partial u}\left[\frac{H(E_v - F_u - H\theta_u) + F(H_u + E\theta_v - F\theta_u)}{EH}\right].$$

The last equation may be reduced to

$$\Delta_2\theta = \frac{1}{H}\frac{\partial}{\partial u}\left[\frac{\partial}{\partial v}\left(\log\frac{E}{H}\right) - \frac{H}{E}\frac{\partial}{\partial u}\left(\frac{F}{H}\right)\right]. \tag{6}$$

If θ is a solution of (6), ρ is found from (5) by a quadrature, then y from (4) by a quadrature. The equation $y = c$ gives an isothermal family and is the general solution of $dy = 0$ or

$$\frac{dv}{du} = \frac{E}{H\cot\theta - F}.$$

Comparing the last equation with (1) it appears that the necessary and sufficient condition that the differential equation, $dv/du = \alpha$, define an isothermal family is that the angle φ measured from (v) to the integral curve C and equal to

$$\text{arc tan}\,\frac{H\alpha}{E + F\alpha}$$

be a solution θ of (6). The angles of the complex integrating factors of the two differential equations of the minimal lines (2) are plus and minus the angle of intersection of the curves of an isothermal family with the curves (v).

When u, v are isothermic parameters, the condition given is that of Lie, for equations (1) and (6) become

$$\tan\varphi = \alpha, \qquad \frac{\partial^2\theta}{\partial u^2} + \frac{\partial^2\theta}{\partial v^2} = 0.$$

We remark that if θ_1 and θ_2 are two solutions of (6), then $\Delta_2(\theta_1 - \theta_2) = 0$, that is, the angle of intersection ψ of two isothermal families is such that

$\Delta_2\psi = 0$, in particular a harmonic function of any pair of isothermic parameters. It may also be easily proved that when u, v are isothermic parameters the necessary and sufficient condition that the equation,

$$du^2 - dv^2 + 2 \tan \varphi du dv = 0,$$

define an isothermal system is that φ be harmonic.

New Haven,
December 3, 1920.

THE REVERSION OF CLASS NUMBER RELATIONS AND THE TOTAL REPRESENTATION OF INTEGERS AS SUMS OF SQUARES OR TRIANGULAR NUMBERS.

By E. T. Bell.

We shall discuss a set of new arithmetical functions defined in §§ 7, 8 relating to representations as sums of square or triangular numbers, the connection of these with class numbers, and means for calculating by recurrence the numerical values of the functions. The new functions first present themselves in reversing the class number formulas of the classical types due to Kronecker, Hermite and Liouville. They are themselves connected by many relations of a like simplicity, and seem to deserve attention on their own account. In section I we fix the notation and state the sense in which reversion is used throughout; in II the functions are defined and their generating series determined, the absolute convergence of these being proved incidentally; III contains four examples of the reversion of simple class number relations, and IV gives a short selection from the numerous recurrences between the functions, those chosen for presentation being among the most useful for numerical computations.

I. Notation; Reversions.

In the customary notation let $F(n)$, $F_1(n)$ denote the number of odd, of even classes respectively of binary quadratic forms for the determinant $-n$, so that $G(n) = F_1(n) + F(n)$ is the whole number of classes, and write

$$H(n) = F(n) - F_1(n).$$

By the usual conventions a class equivalent to $a(x^2 + y^2)$ contributes $\frac{1}{2}$ to F or F_1; one equivalent to $a(2x^2 + 2xy + 2y^2)$ counts for $1/3$ in F_1. It is simpler in the sequel to ignore the other conventions $F(0) = 0$, $F_1(0) = -1/12$; hence all formulas involving $F(n)$, $F_1(n)$ or other arithmetical functions will be so stated as to preclude the occurrence of zero values of the argument n.

Henceforth, without further references, $(\alpha | \beta)$ is the Jacobi-Legendre symbol; m, μ, n, a, b, t are integers > 0, of which m, μ are odd, n, a, b arbitrary, t is triangular $(= 1, 3, 6, 10, \cdots)$, and k is an integer $\geqq 0$. In all power series in q, in particular in those for the elliptic theta constants $\vartheta_a^{\prime} = \vartheta_a(q)$, $\vartheta_1' = \vartheta_0\vartheta_2\vartheta_3$, $\vartheta_0 = 1 + 2\Sigma(-1)^n q^{n^2}$, $\vartheta_2(q^4) = 2\Sigma q^{m^2}$, $\vartheta_3 = 1$

$+ 2\Sigma q^{n^2}$, $\vartheta_1'(q^4) = 2\Sigma(-1 \mid m) m q^{m^2}$, the summations refer to all values from 1 to ∞ of the exponents consistent with the m, n notation. But in all sums independent of q, such as $\Sigma f(2m - a^2)$, the Σ is with respect to the letters μ, a, b, or t involved, and extends only to all those values (of the μ, a, b, or t) that make the argument > 0, so that any such sum consists of only a finite number of terms and zero values of the argument do not occur. When in any sum any of the integers are restricted beyond the notation already explained the restrictions will be given explicitly. Thus

$$m = 8k + 3 : \Sigma f(2m - a^2) = 0$$

indicates that the sum vanishes only when $m \equiv 3 \bmod 8$.

2. A function $f(x)$ which takes a single definite value when x is an integer > 0 is called arithmetical. Let α, β denote arithmetical functions between which there is the relation

$$(1) \qquad \alpha(n) = (-1)^n \begin{vmatrix} \beta(1) & \beta(2) & \beta(3) & \cdots & \beta(n) \\ 1 & \beta(1) & \beta(2) & \cdots & \beta(n-1) \\ 0 & 1 & \beta(1) & \cdots & \beta(n-2) \\ 0 & 0 & 1 & \cdots & \beta(n-3) \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & \beta(2) \\ 0 & 0 & 0 & \cdots & \beta(1) \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix}.$$

We shall call $\alpha(n)$ the inverse of $\beta(n)$, or simply α the inverse of β, a reason for this nomenclature appearing in a moment. On expanding the determinant by minors of the elements in its last column, we see that (1) is equivalent to

$$(2) \qquad \alpha(n) + \beta(n) + \Sigma\alpha(a)\beta(n - a) = 0;$$

and this being symmetric in α, β, it follows that if α is the inverse of β, then β is the inverse of α.

3. The problem of reversing class number relations is presently reduced to finding the inverses of certain elementary arithmetical functions. When α is arithmetically defined it is not always easy à priori to give an explicit arithmetical definition of its inverse β. Thus if $H'(n)$ is the inverse of $12H(n)$ defined in § 1, it may be verified from Dirichlet's formulas for the class number combined with Gauss' theorems on decompositions into sums of three squares that

$$H'(n) = \sum_{r=1}^{n}(-1)^r t_{r+1} N_r'(n),$$

in which t_a denotes the ath triangular number and $N_r'(n)$ is the total number of representations of n as a sum of r squares whose roots are $\geqq 0$. But the verification is artificial and involved, and all such questions are better treated by direct algebraic methods, one of which is elaborated in this paper in detail sufficient for the reversions of class number formulas of any of the classical types.

4. Consider three pairs of arithmetical functions, (P, P'), (Q, Q'), (R, R'), the functions in any pair being inverses of each other, and let P, Q, R be connected by the relation

$$P(n) + Q(n) + \Sigma P(a)Q(n-a) = R(n), \tag{3}$$

which may conveniently be symbolized by PQR in which, note, the function given explicitly in terms of the other two occurs last. The process of solving PQR for P is called the reversion of PQR with respect to P, and the solution the P-reverse of (3). We will show that

$$R(n) + Q'(n) + \Sigma Q'(a)R(n-a) = P(n); \tag{4}$$

that is, the P-reverse of PQR is $RQ'P$, or what is the same thing, by symmetry, $Q'RP$. Hence we have the rule: To reverse any relation of the form PQR with respect to either function given implicitly, interchange the function given explicitly and the function with respect to which the reversion is taken, and replace the other function by its inverse.

It is easily seen that (3) implies (4), viz., that PQR implies $RQ'P$. For if in (4) we replace $R(n)$, $R(n-a)$ by their values as given by (3), the latter being

$$R(n-a) = P(n-a) + Q(n-a) + \sum_b P(b)Q(n-a-b),$$

and collect coefficients of $P(a)$, we find

$$\begin{aligned}[Q(n) &+ Q'(n) + \Sigma Q'(a)Q(n-a)] \\ &+ \sum_a [Q(n-a) + Q'(n-a) + \sum_b Q'(b)Q(n-a-b)]P(a) = 0,\end{aligned}$$

which is an identity, each square bracket vanishing separately since Q, Q' are inverses.

We have just shown that PQR implies $RQ'P$. From this it follows, since if Q' is the inverse of Q then Q is the inverse of Q', that $RQ'P$ implies PQR. Hence in the meaning of mathematical logic PQR and $RQ'P$ are formally equivalent, $PQR \equiv RQ'P$.

It is evident that from any relation of the type PQR we can by reversions obtain six and only six relations of the same type,

$$Q'RP, \quad P'RQ, \quad PQR, \quad QR'P', \quad PR'Q', \quad P'Q'R'.$$

These determine, in the same order, P, Q, R, P', Q', R' from the given relation PQR; and clearly from what precedes, any two of the six are formally equivalent and each implies all.

A relation of type PQR is thus six-valued, and the six values constitute its complete reversion. When considering the reversion of class number relations we shall confine the discussion to the partial reversions which give the class number functions explicitly in terms of known functions and their inverses.

5. We need also the reverses of another type of relation, (PQR), viz.,

$$P(n) + \Sigma P(a)Q(n - a) = R(n). \tag{6}$$

As before it is seen at once that this is two-valued, and the complete reversion is

$$(PQR), \qquad (RQ'P). \tag{7}$$

6. If for all values of q defined by $0 < |q| < c$ where c is a constant, the series

$$\gamma(f) = 1 + \Sigma q^n f(n)$$

converges absolutely, $\gamma(f)$ is called the generator of f. Let f, f' be inverses, and suppose that for the same q both $\gamma(f)$ and $\gamma(f')$ are absolutely convergent. Then from (2) we have, on collecting coefficients of q^n,

$$\gamma(f)\gamma(f') = 1.$$

Hence if for the same q the generators of a function and its inverse are absolutely convergent, the generator of the inverse is the reciprocal of the generator of the function.

Suppose $\gamma(P)$, $\gamma(P')$, $\gamma(Q)$, $\gamma(Q')$, $\gamma(R)$, $\gamma(R')$ are absolutely convergent for the same q, the functions being those in § 4, and suppose further that $\gamma(P)\gamma(Q) = \gamma(R)$. On equating coefficients of q^n in this we find PQR of § 4. Multiplying the identity between the generators throughout by $\gamma(Q')$ we get $\gamma(R)\gamma(Q') = \gamma(P)$, which yields the relation $RQ'P$, viz., the P-reverse of PQR. In this way we find by the appropriate multiplications all six of the relations in the complete reversion of PQR, and similarly for (PQR).

As all of the generators giving rise to class number relations, likewise all of those for the inverses of the several functions occurring in these are absolutely convergent for the same q (see § 9), we shall use the method of generators exclusively in finding the reversions. This method, when it can be applied, is preferable to a direct use of (4), (7) as even in simple cases the necessary arithmetical reductions for the latter are not always apparent.

II. Total Functions and Their Generators.

7. The functions defined in § 8 may be regarded as a natural extension of certain functions occurring in the theory of partitions, as they relate to the total number of ways in which an integer may be written as a sum of square or triangular numbers of preassigned forms. By the total number of ways in which n may be written as a sum of squares we mean the sum of the number of ways in which n may be represented as a sum of r squares whose roots are $\geqq 0$, for $r = 1, 2, \cdots, n$, the order of the squares in any representation being essential. Similarly for the other total functions; *all* the representations of the kinds specified are to be counted, and in each case only squares whose roots are different from zero, or positive triangular numbers, are enumerated in any representation.

It will be noticed in the following functions that the suffix 1 or 2 is of the same parity as the total numbers of odd squares occurring in the several representations of the kinds even (E), or odd (O, Ω), with a similar device for the triangular T, so that the meanings and elementary properties of all the symbols are easily retained.

8. Let $E_1(n)$, $E_2(n)$, $\cdots$, denote the total numbers of representations of n as sums of the following kinds:

(8) $E_1(n)$: even number of squares, the total number of odd squares in each of the representations enumerated being odd; $E_1(2n) = 0$.

(9) $E_2(n)$: even number of squares, the total number of odd squares in each of the representations enumerated being even; $E_2(m) = 0$.

(10) $E(n)$: even number of squares; $E(n) = E_1(n) + E_2(n)$.

(11) $O_1(n)$: odd number of squares, the total number of odd squares in each of the representations enumerated being odd; $O_1(2n) = 0$.

(12) $O_2(n)$: odd number of squares, the total number of odd squares in each of the representations enumerated being even; $O_2(m) = 0$.

(13) $O(n)$: odd number of squares; $O(n) = O_1(n) + O_2(n)$.

(14) $N(n)$: squares; $N(n) = E(n) + O(n)$.

(15) $\Omega_1(n)$: odd number of odd squares; $\Omega_1(2n) = 0$.

(16) $\Omega_2(n)$: even number of odd squares; $\Omega_2(m) = 0$.

(17) $\Omega(n)$: odd squares; $\Omega(n) = \Omega_1(n) + \Omega_2(n)$.

(18) $T_1(n)$: odd number of triangular numbers.

(19) $T_2(n)$: even number of triangular numbers.

(20) $T(n)$: triangular numbers; $T(n) = T_1(n) + T_2(n)$.

(21) $\Phi'(n) = \Phi_2(n) - \Phi_1(n)$, $\Phi = E, O, \Omega, T$.

(22) $D(n) = E(n) - O(n)$; $D'(n) = E'(n) - O'(n)$.

Since $E_1(2n) = 0$, $E_2(2n) = E(2n)$, etc., it may seem that E_1, E_2, O_1, O_2, Ω_1, Ω_2 are superfluous, and that E, O, Ω only are necessary. This of course is true. Nevertheless the statement of many processes is much

simplified by retaining all, which we shall do, using one set or the other as convenient. From the definitions we have the useful identities

(23) $$(-1)^n\Psi(n) = \Psi'(n), \qquad \Psi = E, O, \Omega, D.$$

The functions (8)–(22) are those most frequently required in class number reversions.

9. Let φ denote any one of the functions defined in (8)–(22) except those involving T, T_1 or T_2. Then obviously $|N(n)| \geqq |\varphi(n)|$. Hence the absolute convergency of $\gamma(N)$ for $0 < |q| < c$ implies that of $\gamma(\varphi)$ for the same q. We shall prove that $c = \frac{1}{4}$ ensures the absolute convergency of $\gamma(N)$. This value also makes each of ϑ_a, ϑ_1', and hence also their positive integral powers $\vartheta_a{}^a$ and products $\vartheta_a{}^a\vartheta_\beta{}^b$, etc., absolutely convergent. A larger c ($= 1$) may be found making $\gamma(N)$ absolutely convergent; but as this is needed in nothing that follows, and as the proof is longer, we omit consideration of this point.

Let $N_r(n)$ denote the total number of representations of n as a sum of r squares whose roots are > 0, and $N_r'(n)$ the total number of representations of n as a sum of r squares whose roots are $\geqq 0$. Then $N_r'(n) = 2^rN_r(n)$;

$$N(n) = \sum_{r=1}^{n} N_r'(n) = \sum_{r=1}^{n} 2^r N_r(n) \equiv 2^n \sum_{r=1}^{n} N_r(n).$$

Hence $N(n)/2^n \equiv$ the total number of ways in which n may be written as a sum of squares whose roots are > 0. But clearly this last number $\equiv 2^{n-1}$, for 2^{n-1} is the total number of ways into which n may be partitioned into n or fewer positive non-zero integers. Hence $N(n) \equiv 2^{2n-1}$; and therefore since $1 + \Sigma q^n 2^{2n-1}$ converges absolutely if $0 < |q| < \frac{1}{4}$, the absolute convergence of $\gamma(\varphi)$ for the same range is established. And it is obvious that by a few slight changes this argument can be modified to fit those functions of (18)–(20) which involve T, T_1, T_2. Henceforth this value of q is assumed in all the series.

10. Expanding $\vartheta_3{}^{-1}$ in powers of $\vartheta_3 - 1$,

$$\vartheta_3{}^{-1} = [1 + (\vartheta_3 - 1)]^{-1} = 1 + \Sigma(-1)^n(\vartheta_3 - 1)^n,$$

we see at once that the coefficient of q^n is $D(n)$. Thus we have the first fundamental generator,

(24) $$\gamma(D) = 1/\vartheta_3.$$

Change q into $-q$ and apply (23):

(25) $$\gamma(D') = 1/\vartheta_0.$$

Similarly from $[1 - (\vartheta_3 - 1)]^{-1}$ we derive the second fundamental series

(26) $$\gamma(N) = 1/(2 - \vartheta_3);$$

whence, as before,

$$1 + \Sigma q^n[E'(n) + O'(n)] = 1/(2 - \vartheta_0). \tag{27}$$

From (24), (26) by addition and subtraction,

$$\gamma(E) = 1/(2\vartheta_3 - \vartheta_3^2), \tag{28}$$
$$\gamma(O) - 1 = (\vartheta_3 - 1)/(2\vartheta_3 - \vartheta_3^2); \tag{29}$$

and from these on replacing q by $-q$, or independently from (25), (27),

$$\gamma(E') = 1/(2\vartheta_0 - \vartheta_0^2), \tag{30}$$
$$\gamma(O') - 1 = (\vartheta_0 - 1)/(2\vartheta_0 - \vartheta_0^2). \tag{31}$$

By combining (28)–(31) by addition and subtraction we find $\gamma(\Phi) - 1$ for $\Phi = E_1, E_2, O_1, O_2$. The results, which reduce to comparatively simple forms on factoring numerators and denominators, need not be written out here, as they are required in nothing that follows.

The third fundamental series generates T'; and as before the following are readily seen:

$$1 + \Sigma q^{8n} T'(n) = 2q/\vartheta_2(q^4); \tag{32}$$
$$1 + \Sigma q^{8n} T(n) = 2q/[4q - \vartheta_2(q^4)]; \tag{33}$$
$$1 + \Sigma q^{8n} T_2(n) = 4q^2/\vartheta_2(q^4)[4q - \vartheta_2(q^4)]; \tag{34}$$
$$\Sigma q^{8n} T_1(n) = q[2\vartheta_2(q^4) - 4q]/\vartheta_2(q^4)[4q - \vartheta_2(q^4)]; \tag{35}$$

whence, replacing q by $\sqrt[8]{q}$ we have at once

$$\gamma(T'), \qquad \gamma(T), \qquad \gamma(T_2), \qquad \gamma(T_1) - 1.$$

The fourth set is for Ω:

$$1 + \Sigma q^n \Omega(n) = 1/[1 - \vartheta_2(q^4)]; \tag{36}$$
$$1 + \Sigma q^n \Omega'(n) = 1/[1 + \vartheta_2(q^4)]; \tag{37}$$
$$1 + \Sigma q^n \Omega_2(2n) = 1/[1 - \vartheta_2^2(q^2)]; \tag{38}$$
$$\Sigma q^m \Omega_1(m) = \vartheta_2(q^4)/[1 - \vartheta_2^2(q^4)]. \tag{39}$$

For convenience in numerical checks there is a short table at the end of the paper. All formulas from now on have been checked by means of the table, which was calculated independently. Note that our class number functions will not all agree with those read off from Cayley's table,* as he does not adopt the conventions of § 1. The values in this paper are those which will check in the class number relations of Kronecker, Hermite, Liouville and Humbert, the final form of Kronecker's being that followed.

* Collected Papers, vol. 5, p. 141.

III. Reversions of Class Number Recurrences.

11. The range of possibilities being very extensive we shall discuss only the four class number relations arising from Hermite's (or Kronecker's) developments* of ϑ_3^3, ϑ_2^3, $\vartheta_2\vartheta_3^2$, $\vartheta_2^2\vartheta_3$.

To state the relations we require the functions ζ, ζ', e, λ: $\zeta(n)$ = the sum of all the divisors of n; $\zeta'(n)$ = the sum of the odd divisors of n, $\zeta'(m) = \zeta(m)$; $e(n) = 1$ or 0 according as n is or is not the square of an integer > 0;

$$\lambda(n) = [1 + 2(-1)^n]\zeta'(n); \quad \therefore \lambda(m) = -\zeta(m), \quad \lambda(2n) = 3\zeta'(n).$$

From the developments of the elliptic constants in the Fundamenta Nova, or from the theorems on representations of integers as sums of four squares, we have

$$(40) \qquad \vartheta_0^4 = 1 + 8\Sigma q^n\lambda(n), \qquad \vartheta_2^4 = 16\Sigma q^m\zeta(m);$$

$$(41) \qquad \vartheta_2^2(q^2)\vartheta_3^2(q^2) = 4\Sigma q^m\zeta(m), \qquad \vartheta_0^2\vartheta_3^2 = 1 + 8\Sigma q^{2n}\lambda(n);$$

and Hermite's series are

$$(42) \qquad \vartheta_3^3 = 1 + 12\Sigma q^n H(n), \qquad \vartheta_3(q^2)\vartheta_2^2(q^2) = 4\Sigma q^m F(2m);$$

$$(43) \qquad m = 4k + 1: \qquad \vartheta_2(q^4)\vartheta_3^2(q^4) = 4\Sigma q^m F(m);$$

$$(44) \qquad m = 8k + 3: \qquad \vartheta_2^3(q^4) = 8\Sigma q^m F(m).$$

12. Using the series in (42), (40), the last after replacing q by $-q$, we find in the usual way from the identity $\vartheta_3 \times \vartheta_3^3 = \vartheta_3^4$ the class number relation

$$(45) \qquad 6H(n) + 12\Sigma H(n - a^2) = 4(-1)^n\lambda(n) - e(n).$$

To reverse this we proceed as in § 6, finding the H-reverse from the identity $\vartheta_3^3 = \vartheta_3^4 \times 1/\vartheta_3$ by means of (24), getting at once

$$(46) \qquad 12H(n) = 8(-1)^n\lambda(n) + D(n) + 8\Sigma(-1)^a\lambda(a)D(n - a).$$

13. Similarly from $\vartheta_2^3 \times \vartheta_2 = \vartheta_2^4$ we find by (44) the relation

$$(47) \qquad \Sigma F(4m - \mu^2) = \zeta(m);$$

and from $\vartheta_2^3 = \vartheta_2^4 \times 1/\vartheta_2$ by (32) the F-reverse of this,

$$(48) \qquad F(4m - 1) = \zeta(m) + \Sigma\zeta(\mu)T'\left(\frac{m - \mu}{2}\right),$$

which, on making an obvious change in notation, may be written more conveniently,

$$(49) \qquad m = 8k + 3: \quad F(m) = \zeta\left(\frac{m + 1}{4}\right) + \Sigma\zeta(\mu)T'\left(\frac{m - 4\mu + 1}{8}\right).$$

* Hermite, J. des Math., 1862, p. 25, and formulas (A), (B), (C) of Oeuvres, vol. 4, p. 138. The right of (C) is a misprint for $\vartheta_2^3(q)$. Notice his convention regarding F in formula (B); see footnote to § 14.

14. From $\vartheta_2\vartheta_3^2 \times \vartheta_2 = \vartheta_2^2\vartheta_3^2$ we have in the same way by (43), (41)

$$(50) \qquad 2\Sigma F(2m - \mu^2) = \zeta(m);$$

and from $2\sqrt[4]{q}\vartheta_2\vartheta_3^2 = \vartheta_2^2\vartheta_3^2 \times 2\sqrt[4]{q}/\vartheta_2$ on using (32), we find the F-reverse

$$(51) \quad 2F(2m-1) = \zeta(m) + \tfrac{1}{2}\Sigma[1 + (-1 \mid \mu m)]\zeta(\mu)T'\left(\frac{m-\mu}{4}\right).$$

The factor $\frac{1}{2}[1 + (-1 \mid \mu m)] = 1$ or 0 according as μ, m are congruent or incongruent modulo 4. In (50), (51), as always, F is taken with the usual conventions* for Kronecker's formulas.

15. As a last example we find the F-reverse of

$$(52) \qquad F(2m) + 2\Sigma F(2m - 4a^2) = \zeta(m)$$

which comes from $\vartheta_2^2\vartheta_3 \times \vartheta_3 = \vartheta_2^2\vartheta_3^2$;

$$(53) \qquad F(2m) = \zeta(m) + \Sigma\zeta(\mu)D\left(\frac{m-\mu}{2}\right).$$

By § 4 each of the pairs (45) and (46), (47) and (49), (50) and (51), (52) and (53) are formally equivalent in the sense that each member of a pair implies the other, and it is possible to transform each into the other arithmetically. Again, the first member in any pair is a reverse of the second member with respect to a certain function; thus (52) is the e_2-reverse of (53), where $e_2(n) = 1$ or 0 according as n is or is not the square of an even integer > 0. Each of the pairs may be reduced to several different forms by means of the elementary properties of F, F_1, G, H.

The other relations of the classical types involve what Hermite called incomplete functions instead of the complete functions ζ, ζ', λ, viz., functions of the divisors d, δ of n subject to inequalities, such as $d < \delta$. The reversion of such relations requires the definitions of several (incomplete) functions, but introduces no principle distinct from the preceding.

IV. Recurrences for the Total Functions.

16. To state these we require but one more well-known function, $\xi(n)$, = the excess of the number of divisors of n that are $\equiv 1 \bmod 4$ over the number $\equiv 3 \bmod 4$, so that $\xi(4k+3) = 0$, and $4\xi(n) =$ the number of representations of n as a sum of two squares whose roots are $\gtreqless 0$. The recurrences are derived by the same simple process as the class number

* This is emphasized because there seems to be some confusion in Hermite's notation for the development of $\vartheta_2\vartheta_3^2$ (Oeuvres, vol. 4, pp. 138, 148). We must take $F(1)$, $F(9)$, $F(25)$, $F(49)$, $F(81)$, $\cdots = 1/2, 5/2, 5/2, 9/2, 17/2, \cdots$ in accord with the usual conventions and not, as Hermite appears to intend, $0, 2, 2, 4, 8, \cdots$. This may be verified by putting $m = 1, 5, 13, 25, 41, \cdots$ in (50).

relations and their reversions in the preceding section. Some are obvious from the definitions, others are less evident.

17. Multiplying both sides of (24) by ϑ_3 and equating coefficients of q^n, we get

$$D(n) + 2\Sigma D(n - a^2) = -2e(n); \tag{54}$$

and from (26) in the same way

$$N(n) - 2\Sigma N(n - a^2) = 2e(n); \tag{55}$$

whence, by adding and subtracting,

$$E(n) = 2\Sigma O(n - a^2), \qquad O(n) = 2\Sigma E(n - a^2) + 2e(n), \tag{56}$$

from which with the initial condition $E(1) = 0$, O, E, and hence O_1, O_2, E_1, E_2 may be rapidly calculated. It is advantageous in practice to separate the cases. Replacing E, O by $E_1 + E_2$, $O_1 + O_2$ respectively in (62), and combining, we have

$$\begin{aligned} E_1(m) &= 2\Sigma[O_1(m - 4a^2) + O_2(m - \mu^2)], \\ E_2(2n) &= 2\Sigma[O_1(2n - \mu^2) + O_2(2n - 4a^2)], \\ O_1(m) &= 2\Sigma[E_1(m - 4a^2) + E_2(m - \mu^2)] + 2e(m), \\ O_2(m) &= 2\Sigma[E_1(2n - \mu^2) + E_2(2n - 4a^2)] + 2e(2n), \end{aligned} \tag{57}$$

which, with the initial conditions

$$E_1(1) = 0, \qquad E_2(2) = 4, \qquad O_1(1) = 2, \qquad O_2(2) = 0,$$

suffice for the simultaneous computation by recurrence of the four functions. Obviously the suffixes in (57) may be suppressed.

18. We may eliminate O, E in turn from (56), getting thus recurrences involving O, E separately. On reduction of the results by means of the theorems for the representations of a number as a sum of two squares, these recurrences may be cast into forms involving single summations in place of the double introduced by the elimination. It is simpler, however, to derive these otherwise. We have $\vartheta_3^2 = 1 + 4\Sigma q^n \xi(n)$; hence from (28), proceeding as in § 17, we find

$$E(n) - 4\Sigma[\xi(a) - e(a)]E(n - a) = 4[\xi(n) - e(n)]; \tag{58}$$

and similarly from (29),

$$O(n) - 4\Sigma[\xi(a) - e(a)]O(n - a) = 2e(n); \tag{59}$$

whence, adding and subtracting, we have

$$N(n) - 4\Sigma[\xi(a) - e(a)]N(n - a) = 2[2\xi(n) - e(n)], \tag{60}$$
$$D(n) - 4\Sigma[\xi(a) - e(a)]D(n - a) = 2[2\xi(n) - 3e(n)]. \tag{61}$$

In using these we require the successive values of $\xi(n)$, which presuppose the resolution of $1, 2, \cdots, n$ into prime factors. To avoid this tentative process we find a recurrence for the computation of $\xi(n)$ from the identity $\vartheta_0 \times \vartheta_2\vartheta_3 = \vartheta_1'$:

$$(62) \quad m = 4k + 1: \quad \xi(m) + 2\Sigma(-1)^a\xi(m - 4a^2) = e(m)\sqrt{m}(-1 \mid \sqrt{m}),$$

which, with $\xi(4k + 3) = 0$ and $\xi(2^a m) = \xi(m)$, is sufficient for the non-tentative calculation of all the coefficients in (58)–(61). There is another recurrence for $\xi(n)$, but it is less simple than (62). Incidentally we note that (62) enables us to calculate the number of representations of an integer as a sum of two squares by recurrence. There are similar theorems for any *odd* number of squares up to 13.*

19. Passing to recurrences for the T functions we have from (32) on multiplying throughout by $\vartheta_2(q^4)$, and equating coefficients of like powers of q,

$$m = 8k + 1: \qquad \Sigma T'\left(\frac{m - \mu^2}{8}\right) = -e(m);$$

whence, by an obvious change in notation,

$$(63) \qquad T'(n) + \Sigma T'\left(\frac{8n + 1 - (\mu + 2)^2}{8}\right) = -e(8n + 1).$$

In the same way from (33),

$$(64) \qquad T(n) - \Sigma T\left(\frac{8n + 1 - (\mu + 2)^2}{8}\right) = e(8n + 1),$$

and combining (63), (64) we have the following for the calculation simultaneously of T_1, T_2,

$$(65) \qquad T_1(n) = e(8n + 1) + \Sigma T_2\left(\frac{8n + 1 - (\mu + 2)^2}{8}\right),$$

$$(66) \qquad T_2(n) = \Sigma T_1\left(\frac{8n + 1 - (\mu + 2)^2}{8}\right).$$

From (34), (35) we find recurrences for the separate calculation of T_1, T_2:

$$(67) \quad T_1(n) = e(8n + 1) + \Sigma[\xi(4a + 1) - 2e(8a + 1)]T_1(n - a),$$

$$(68) \quad \begin{aligned} T_2(n) &= \xi(4n + 1) - 2e(8n + 1) \\ &\quad + \Sigma[\xi(4a + 1) - 2e(8a + 1)]T_2(n - a); \end{aligned}$$

whence, by addition and subtraction,

$$(69) \quad \begin{aligned} T(n) &= \xi(4n + 1) - e(8n + 1) \\ &\quad + \Sigma[\xi(4a + 1) - 2e(8a + 1)]T(n - a), \end{aligned}$$

$$(70) \quad \begin{aligned} T'(n) &= \xi(4n + 1) - 3e(8n + 1) \\ &\quad + \Sigma[\xi(4a + 1) - 2e(8a + 1)]T'(n - a). \end{aligned}$$

* American Journal, July, 1920.

20. From (36)–(39) we similarly derive the following for the Ω-functions:

(71) $$\Omega(n) = [1 - (-1)^n]e(n) + 2\Sigma\Omega(n - \mu^2),$$

(72) $$\Omega'(n) = [(-1)^n - 1]e(n) - 2\Sigma\Omega'(n - \mu^2),$$

(73) $$\Omega_2(2n) = 2[1 - (-1)^n]\xi(n) + 4\Sigma\xi(\mu)\Omega_2(2n - 2\mu),$$

(74) $$\Omega_1(m) = 2e(m) + 4\Sigma\xi(\mu)\Omega_1(m - 2\mu).$$

21. The appended table with $F(11) = 3$, $F(19) = 3$, $F(27) = 4$, will be found sufficient for the numerical verification of all formulas in sections III, IV. In using the table we make the elementary transformations $E_2(2n) = E(2n)$, $N(n) = E(n) + O(n)$, etc., whenever necessary.

n	E	O	T_1	T_2	Ω	F	H	ξ	ζ	λ
1	0	2	1	0	2	1/2	1/2	1	1	− 1
2	4	0	0	1	4	1	1	1	3	3
3	0	8	2	0	8	1	2/3	0	4	− 4
4	16	2	0	3	16	1	$\frac{1}{2}$	1	7	3
5	8	32	4	0	32	2	2	2	6	− 6
6	64	24	1	6	64	2	2	0	12	12
7	64	128	9	2	128	1	0	0	8	− 8
8	260	160	3	13	256	2	1	1	15	3
9	384	538	19	6	514	5/2	5/2	1	13	−13
10	1128	896	12	28	1032	2	2	2	18	18

UNIVERSITY OF WASHINGTON.

NOTE ON THE TERM MAXIMAL SUBGROUP.

By G. A. Miller.

The term *maximal subgroup*, or maximum subgroup, is the source of so much confusion on the part of the student of group theory that it seems worth while to consider the feasibility of replacing it by some other term. As such a term we would suggest *primary subgroup*. Instead of saying that the subgroup composed of all the substitions of a primitive group which omit a letter is maximal we should then say that this subgroup is primary, and thus associate the terms primary and primitive. Even if such a change of terms should not appear feasible a consideration of the objectional features of the term maximal subgroup may tend to reduce the confusion due to its use. This confusion is the more regretable because of the fact that it relates to elementary and fundamental properties of groups.

The size of a finite group is commonly measured by its order. If two groups have different orders, the one which has the larger order is said to be the larger group. This method of determining the relative magnitudes is also commonly used as regards subgroups. On the other hand, it is customary to call a subgroup a maximal subgroup, or a largest subgroup, even when the group contains subgroups whose orders are larger than that of this maximal subgroup. A necessary and sufficient condition that a subgroup is maximal is that it is not contained in a larger subgroup. In particular, the icosahedral group contains maximal subgroups of each of the following orders: 6, 10, 12.

It is possible to find a series of subgroups of any group G, beginning with any maximal subgroup G_1 and ending with the identity

$$G_1, G_2, \cdots, G_\lambda = 1,$$

such that the smallest subgroup of G which contains any of these subgroups besides G_1 is the one which precedes it in this series, while G_1 is not contained in any subgroup of G. The position of G_1 in this series seems to justify the term primary subgroup as a suggestive term for it. The subgroup which immediately precedes the identity in this series is of prime order. When the order of G is p^m, p being a prime number, $\lambda = m$.

In a group of prime power order every maximal subgroup is also a subgroup of maximal order and every maximal invariant subgroup is also of maximal order. Hence it might at first appear that the use of the term maximal subgroup as regards these groups would be unobjectionable.

That this is not the case results directly from the use of the term maximal abelian invariant subgroup.* If such a subgroup of a non-abelian group of order p^m is of order p^α, this group may contain larger invariant abelian subgroups as may be seen from the group of order 2^{10} defined as follows:

Let $s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8$ and $s_1, s_2, s_3, s_4, s_5, s_6, t_7, t_8$ be sets of generators of two abelian groups of order 2^8 and of type $(1, 1, 1, \cdots)$ and suppose that

$$t_7s_7t_7 = s_1s_7, \qquad t_7s_8t_7 = s_2s_8, \qquad t_8s_7t_8 = s_3s_7, \qquad t_8s_8t_8 = s_4s_8.$$

The group of order 2^{10} generated by the ten operators $s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, t_7, t_8$ has the interesting property that it contains two and only two abelian subgroups of order 2^8. A similar group can easily be constructed for every value of p and each of the groups thus constructed contains two and only two abelian subgroups of order p^8. These two subgroups are evidently both invariant and maximal abelian subgroups. They illustrate a statement made without proof in the *Finite Groups* by Miller, Blichfeldt, Dickson, 1916, page 126.

This group of order 2^{10} is transformed into itself by an operator t_9 which is of order 2 and satisfies the following conditions:

$$t_9s_2t_9 = s_3, \qquad t_9s_5t_9 = s_1s_5, \qquad t_9s_6t_9 = s_4s_6, \qquad t_9s_7t_9 = t_7, \qquad t_9s_8t_9 = t_8.$$

We thus obtain a group of order 2^{11} which has two conjugate abelian subgroups of order 2^8 but no invariant abelian subgroup whose order exceeds 2^7. The abelian subgroup of order 2^7 generated by $s_1, s_2, s_3, s_4, s_5, s_6, s_7, t_7$ is a maximal invariant abelian subgroup of the given group of order 2^{10} notwithstanding the fact that this group contains invariant abelian subgroups of larger order. As similar subgroups exist for all values of p it results that there are groups of order p^m, p being any prime number, which contain larger invariant abelian subgroups than some of their maximal invariant abelian subgroups. Hence it is clear that the term maximal subgroup is apt to lead to confusion even with respect to prime power groups.

It may be of interest to note in this connection that from the known theorem that every abelian subgroup of order p^α which is contained in a group of order p^m is found in $1 + kp$ abelian subgroups of order $p^{\alpha+1}$ whenever it is found in at least one such abelian subgroup† it results directly that every invariant abelian subgroup of order p^α is found in a number of invariant abelian subgroups of order $p^{\alpha+1}$ which is of the form $1 + kp$ whenever it is contained in at least one such subgroup. In particular, every invariant subgroup of any group of order p^m contains a primary or maximal invariant abelian subgroup which is invariant under the entire group. This theorem was proved in a different manner by Burnside in the article to which reference was made.

* Cf. W. Burnside, Proceedings of the London Mathematical Society, vol. 13 (1914), p. 9.

† Miller, Messenger of Mathematics, vol. 36 (1907), p. 70.

REDUCIBLE CUBIC FORMS EXPRESSIBLE RATIONALLY AS DETERMINANTS.

By L. E. Dickson.

1. A quadratic form q in three or four variables can be expressed in general in the form $xy - z^2$ or $xy - zw$, each of which is a determinant of order two. Hence if l is any linear form, lq equals a determinant of order three whose elements are linear functions of the variables.

Henceforth, let l and q have rational coefficients. Can we express lq rationally in determinantal form, i.e., as a determinant whose elements are linear functions with rational coefficients? We cannot ordinarily employ the above special method in which the elements of a row are $l, 0, 0$, since the two linear functions in a row of the minor vanish for rational values, not all zero, of the variables, while q need not vanish for such values. We introduce l as the new variable y. For three variables, yq is always expressible rationally in determinantal form, as shown by taking $w = 0$ in the formula of § 2. For four variables, the question is not so simple, but is answered completely by the following

Theorem. *Let q be a quadratic form in four variables with rational coefficients. (i) If q vanishes at some rational point having $y = 0$, yq is expressible rationally in determinantal form. (ii) If $q \neq 0$ for every rational point having $y \neq 0$, then yq is expressible rationally in determinantal form if and only if either yq is equivalent to a ternary form, or the determinant of q is the square of a rational number $\neq 0$ and the determinant of $q(x, 0, z, w)$ is $\neq 0$. (iii) If both of the preceding hypotheses be denied, so that $q \neq 0$ at every rational point having $y = 0$, and $q = 0$ for some rational point having $y \neq 0$, then yq is not expressible rationally in determinantal form.*

The respective cases are in geometrical language: (*i*) The quadric surface has a rational point in common with the plane. (*ii*) Every rational point of the surface lies in the plane. (*iii*) The surface contains a rational point, but contains no rational point of the plane.

If $q = yL$, yq equals a determinant whose diagonal elements are y, y, L. But if q contains terms free of y, we can apply a linear transformation on x, z, w with rational coefficients which replaces q by a form in which the coefficient of x^2 is $c \neq 0$. We may assume that $c = 1$, since cy may be taken as a new y in yq. After making a suitable addition to x, we obtain yQ, where $Q = x^2 + f$, and f is a quadratic form in y, z, w.

2. First, let Q vanish at a rational point $P = (x', 0, z', w')$ for which $y = 0$. Since z', w' are not both zero, we may take $w' \neq 0$, interchanging z and w if necessary. Taking $zw' - wz'$ as a new variable z, we have $P = (a, 0, 0, 1)$. Let yL denote the sum of the terms of f with the factor y; then

$$Q = x^2 + yL + dz^2 + ezw - a^2w^2, \qquad yQ = \begin{vmatrix} x + aw & y & 0 \\ -L & x - aw & z \\ dz + ew & 0 & y \end{vmatrix}.$$

3. Second, let $Q \neq 0$ for every rational point having $y \neq 0$. Assume that yQ equals a determinant D whose nine elements are linear functions of x, y, z, w with rational coefficients. Since x^2y is the only term involving x in yQ, we may assume that x occurs, with coefficient unity, in the first element of the first row of D and in none of the remaining elements of the first row and first column; also that x occurs, with coefficient unity, in the second element of the second row and not elsewhere in the second row or column; and that the last element of the third row is y. Hence

$$(1) \qquad D = \begin{vmatrix} x + l_1 & l_2 & l_3 \\ l_4 & x - l_1 & l_6 \\ l_7 & l_8 & y \end{vmatrix},$$

where the l's are free of x, while l_3, l_6, l_7, l_8 may be assumed free also of y (in view of the element y), and where the preliminary entry l_5 has been replaced by its value $- l_1$. In fact, the terms linear in x were $xy(l_1 + l_5) - x(l_3l_7 + l_6l_8)$, whence $l_1 + l_5 \equiv 0$ and

$$(2) \qquad l_3l_7 + l_6l_8 \equiv 0.$$

Since $D = 0$ when $x = - l_1$, $l_2 = l_3 = 0$ (which are satisfied by an infinitude of rational values of x, y, z, w), these linear relations must imply $y = 0$, in view of our hypothesis that $yQ \neq 0$ if $y \neq 0$. Hence y equals a linear homogeneous function of l_2 and l_3. But l_3 is free of y. Hence

$$(3) \qquad l_2 \equiv \rho y + \sigma l_3, \qquad \rho \neq 0.$$

Using similarly the elements of the second row, first and second columns, we see that

$$(4) \qquad l_4 \equiv ry + tl_6, \qquad r \neq 0,$$

that l_4 is a linear function of y and l_7, and that l_2 is a linear function of y and l_8. Thus, if $\sigma l_3 \not\equiv 0$,

$$(5) \qquad l_8 = vl_3, \qquad l_7 = - vl_6, \qquad v \neq 0,$$

by (2); the same follow also if $tl_6 \not\equiv 0$. By (5),

$$D = y(x^2 - l_1^2 - l_2l_4) - v\lambda, \qquad \lambda \equiv 2l_1l_3l_6 + l_2l_6^2 - l_3^2l_4.$$

By (3) and (4),

$$\lambda = l_3 l_6(2l_1 + \sigma l_6 - tl_3) + \rho y l_6^2 - r y l_3^2.$$

Since D shall have the factor y, we conclude that

$$l_1 = \tfrac{1}{2}tl_3 - \tfrac{1}{2}\sigma l_6 \tag{6}$$

if $l_3 l_6 \neq 0$ and if l_1 (like l_3 and l_6) is free of y. This is accomplished as follows. Add the products of the elements of the first row of D by k to the elements of the second row, and then subtract the products of the elements of the second column by k from the elements of the first column. We obtain a determinant of the same form as (1) with $l_1 - kl_2$ in place of l_1. By choice of k, we may assume that l_1 lacks y. We now have

$$Q = D/y = x^2 - l_1^2 - l_2 l_4 - v\rho l_6^2 + vrl_3^2. \tag{7}$$

Inserting the values (3), (4), (6) of l_2, l_4, l_1, we obtain a quadratic form in x, y, l_3, l_6, whose determinant equals $(r\rho m)^2$, where

$$m = v - \frac{t^2}{4r} + \frac{\sigma^2}{4\rho},$$

while the determinant of the part in l_3, l_6 only is $-vr\rho m$. Or we may avoid this computation by completing the square of the terms in y and finding that the terms in $l_3 l_6$ cancel:

$$Q = x^2 - r\rho Y^2 + rml_3^2 - \rho ml_6^2, \qquad Y = y + \frac{\sigma}{2\rho} l_3 + \frac{t}{2r} l_6.$$

The determinant of Q is seen by inspection to be $(r\rho m)^2$. If $m = 0$, then $Q = 0$ when $x = Y = 0$, which imply $y = 0$ only when $Y \equiv y$, and then yQ is a binary form.

It remains to consider the special cases excluded above. If $l_3 \equiv l_6 \equiv 0$,

$$Q = D/y = x^2 - l_1^2 - r\rho y^2 \tag{8}$$

vanishes when $x - l_1 = \alpha y$, $x + l_1 = \beta y$, $\alpha\beta = r\rho$, $\alpha \neq \beta$, which imply $y = 0$ only when $l_1 \equiv 0$, and then yQ is a binary form.

If $l_3 \equiv 0$, $l_6 \neq 0$, then $l_5 \equiv 0$ by (2). Using (3) and (4), we get

$$Q = D/y = x^2 - l_1^2 - \rho y(ry + tl_6) + \rho l_6 l_7.$$

Now $Q = 0$ when $x = l_1$, $y = l_6$, $ry + tl_6 = l_7$, which imply $y = 0$ only when l_7 is proportional to l_6. Hence let $l_7 \equiv kl_6$. Then $Q = 0$ when $l_6 = y$, $x - l_1 = \alpha y$, $x + l_1 = \beta y$, $\alpha\beta = \rho(r + t - k)$, $\alpha \neq \beta$, which imply $y = 0$ only when $l_1 \equiv 0$, and then yQ is a ternary form in x, y, l_6.

If $l_3 \neq 0$, $l_6 \equiv 0$, we interchange the first two rows and first two columns of (1) and are led to the preceding case.

Finally, let $l_3 l_6 \neq 0$. By the remarks accompanying (5), it remains to consider only the case in which $\sigma = t = 0$, whence $l_2 = \rho y$, $l_4 = ry$. By (2), l_6 or l_8 is divisible by l_3. In the second alternative, we have (5), since, if $l_8 \equiv 0$, then $l_7 \equiv 0$ and D is of the form (8). Hence $l_6 \equiv sl_3$, $s \neq 0$, so that $l_7 \equiv -sl_8$ by (2). The only term of D lacking y is $-2sl_1l_3l_8$, whence $l_1 l_8 \equiv 0$. But $l_8 \equiv 0$ was seen to lead to (8). Hence $l_1 \equiv 0$ and

$$D/y = x^2 - r\rho y^2 + (r - s^2\rho) l_3 l_8$$

is zero when $x = r$, $y = s$, $l_3 l_8 = -r$. But l_3 and l_8 can be given any desired values by choice of z and w unless they are proportional, which is the case (5) already treated.

Conversely, let the determinant of $x^2 + f$ be a rational square $\neq 0$ and the determinant of $f(0, z, w)$ be $\neq 0$. By a linear transformation altering neither x nor y and having rational coefficients we can evidently delete the terms in zw, yz, yw. We obtain a form of the following type, whose simplest representation as a determinant is obtained by taking $\sigma = t = 0$ in (1)–(7):

$$y(x^2 - r\rho y^2 + rvl_3^2 - \rho v l_6^2) = \begin{vmatrix} x & \rho y & l_3 \\ ry & x & l_6 \\ -vl_6 & vl_3 & y \end{vmatrix}.$$

4. If Q falls under neither § 2 nor § 3, then $Q \neq 0$ for every rational point having $y = 0$, and $Q = 0$ for some rational point P with $y \neq 0$, say $P = (x', 1, z', w')$. We shall prove* that yQ is not equal to a determinant (1) with rational elements. Taking $z - z'y$ and $w - w'y$ as new variables z, w, we may write $P = (a, 1, 0, 0)$. In view of (2), the expansion of (1) gives

$$D = y(x^2 - l_1^2 - l_2 l_4) + l_3 l_4 l_8 + l_2 l_6 l_7 + 2 l_1 l_3 l_7. \tag{9}$$

First, let one of l_2, l_4 contain y. Interchanging rows and columns if necessary, we may assume that

$$l_2 = \rho y + L_2, \qquad l_4 = ry + L_4, \qquad \rho \neq 0,$$

where L_2, L_4 are functions of z, w. By the argument just above (7) in § 3, we may assume that l_1 lacks y. Since (9) shall equal yQ,

$$Q = x^2 - l_1^2 - l_2 l_4 + r l_3 l_8 + \rho l_6 l_7, \qquad R \equiv l_3 l_8 L_4 + l_6 l_7 L_2 + 2 l_1 l_3 l_7 \equiv 0.$$

If $l_3 \equiv l_6 \equiv 0$, $Q = 0$ when $x = l_1$, $y = 0$, $L_2 = 0$, contrary to hypothesis. If $l_3 \equiv 0$, $l_6 \neq 0$, then $l_8 \equiv 0$ by (2), $l_7 L_2 \equiv 0$ by $R \equiv 0$, and $Q = 0$ when $x = l_1$, $y = 0$, $L_2 L_4 = \rho l_6 l_7$, which implies a linear relation between z and w. Hence $l_3 \neq 0$, and, similarly, $l_6 \neq 0$, $l_7 \neq 0$, $l_8 \neq 0$.

* The hypotheses are satisfied if $Q = x^2 - y^2 + 2z^2 + 3w^2$, $P = (1, 1, 0, 0)$.

By (2), l_6 or l_8 is divisible by l_3. In the first case,

$$(10) \qquad l_6 \equiv \alpha l_3, \qquad l_7 \equiv -\alpha l_8, \qquad \alpha \not\equiv 0,$$

whence $2\alpha l_1 \equiv L_4 - \alpha^2 L_2$ by $R \equiv 0$. Then $Q = 0$ for

$$y = 0, \qquad x = \frac{L_4 + \alpha^2 L_2}{2\alpha}, \qquad l_3 = 0,$$

contrary to hypothesis. Hence l_6 and l_3 are not proportional and

$$(11) \qquad l_8 \equiv \beta l_3, \qquad l_7 \equiv -\beta l_6, \qquad \beta \not\equiv 0,$$

whence $l_3^2 L_4 - l_6^2 L_2 - 2l_1 l_3 l_6 \equiv 0$ by $R \equiv 0$. Thus $L_4 \equiv k l_6$ and $k l_3^2 - l_6 L_2 - 2l_1 l_3 \equiv 0$, whence $L_2 \equiv p l_3$, $2l_1 \equiv k l_3 - p l_6$. Thus

$$Q = x^2 - \tfrac{1}{4}(k l_3 - p l_6)^2 - (\rho y + p l_3)(r y + k l_6) + r\beta l_3^2 - \rho\beta l_6^2.$$

Since Q is zero at P, $\rho r = a^2$, whence $r = \rho t^2$, $t = a/\rho$. Thus

$$Q_{y=0} = x^2 - M^2 + \beta\rho(t^2 l_3^2 - l_6^2), \qquad M \equiv \tfrac{1}{2}(k l_3 + p l_6),$$

so that $Q = 0$ for $y = 0$, $x = M$, $l_6 = t l_3$, contrary to hypothesis.

Second, let l_2 and l_4 both lack y. Let $l_1 = cy + L_1$. Then, by (9),

$$Q = x^2 - l_1^2 - l_2 l_4 + 2c l_3 l_7, \qquad S \equiv l_3 l_4 l_8 + l_2 l_6 l_7 + 2L_1 l_3 l_7 \equiv 0.$$

If $l_3 l_7 \equiv 0$, $Q = 0$ when $x = L_1$, $y = 0$, $l_2 = 0$, contrary to hypothesis. Thus $l_3 l_7 \not\equiv 0$ and, similarly, $l_2 l_4 \not\equiv 0$. Then $l_6 l_8 \not\equiv 0$ by (2). We have (10) or (11). By (10), $S \equiv 0$ implies $2\alpha L_1 \equiv l_4 - \alpha^2 l_2$. Then $Q = 0$ when

$$y = 0, \qquad l_3 = 0, \qquad x = (l_4 + \alpha^2 l_2)/(2\alpha).$$

By (11), $S \equiv \beta(l_3^2 l_4 - l_2 l_6^2 - 2L_1 l_3 l_6) \equiv 0$. Hence $l_2 \equiv d l_3$ and then $l_4 \equiv e l_6$, $2L_1 \equiv e l_3 - d l_6$. Thus $Q = 0$ when $y = l_3 = 0$, $x = \frac{1}{2} d l_6$, contrary to hypothesis.

NOTE ON THE PICARD METHOD OF SUCCESSIVE APPROXIMATIONS.

By Dunham Jackson.

The Picard method of successive approximations, as applied to the proof of the existence of a solution of a differential equation of the first order, is commonly introduced somewhat after the following manner:

"We shall develop the method on an equation of the first order

$$(1) \qquad \frac{dy}{dx} = f(x, y),$$

supposing first that the variables are real. We shall assume that the function f is continuous when x varies from x_0 to $x_0 + a$ and when y varies between the limits $(y_0 - b, y_0 + b)$; that the absolute value of the function f remains less than a positive number M when the variables x, y remain within the preceding limits; and, finally, that there exists a positive number A such that we have

$$|f(x, y) - f(x, y')| < A\,|y - y'|$$

for any positions of the points (x, y) and (x, y') in the preceding region.

"Let us suppose, for ease in the reasoning, $a > 0$, and let h be the smaller of the two positive numbers a, b/M. We shall prove that *the equation* (1) *has an integral which is continuous in the interval* $(x_0, x_0 + h)$ *and which takes on the value* y_0 *for* $x = x_0$."

This particular language is quoted substantially from Goursat's Mathematical Analysis, translated by Hedrick and Dunkel,* except that the statement there is for a pair of differential equations in two unknown functions. The italics are kept from the original French. After the proof has been given, the following remark is added:†

"If $\cdots$ we go over the proof again, we see that the condition $h < b/M$ is needed only to make sure that the intermediate functions y_1, y_2, $\cdots$ [the successive approximations to the solution] do not get out of the interval $(y_0 - b, y_0 + b)$, so that the functions $f(x, y_i)$ shall be continuous functions of x between x_0 and $x_0 + h$. If the function $f(x, y)$ remains continuous when x varies from x_0 to $x_0 + a$, and when y varies from $-\infty$ to $+\infty$, it is unnecessary to make this requirement."

* Vol. 2, part 2, pp. 61–62.

† Loc. cit., p. 64; the statement is again simplified from two unknowns to one, in quoting.

The purpose of this note is to point out that even if $f(x, y)$ is originally defined only in a rectangle, it is a simple matter to extend its definition outside the rectangle, so that the conditions of the hypothesis shall hold for $x_0 \leqq x \leqq x_0 + a$ and for all real values of y, the Lipschitz condition as well as the mere continuity. It is sufficient, for example, to let

$$f(x, y) = f(x, y_0 + b), \qquad y \geqq y_0 + b,$$
$$f(x, y) = f(x, y_0 - b), \qquad y \leqq y_0 - b.$$

The process of successive approximations then gives, at a single stroke, a function $y(x)$ which is defined and satisfies the differential equation, with the extended definition of $f(x, y)$, for $x_0 \leqq x \leqq x_0 + a$. It satisfies the *original* equation as long as the x and *y of the solution* remain within the original rectangle, whatever the behavior of the approximating functions may be. The solution is unique as long as it stays in the rectangle. The original equation of course has no authority outside its own domain, and corresponding to the infinitely many possible ways of extending the definition of f there will be infinitely many different extensions of the solution, if it leaves the rectangle before x reaches $x_0 + a$.

It may seem that this observation is trivial, and it is perhaps hardly probable that it is made here for the first time;* but its omission from standard presentations of the subject is notable. In the treatise already quoted, for example, after the Cauchy-Lipschitz proof has been explained, the two demonstrations are compared as follows:†

"Cauchy's first method [the Cauchy-Lipsenitz method] and that of the successive approximations give, as we see, the same limit for the interval in which the integral surely exists. But from a theoretical point of view Cauchy's method is unquestionably superior: we shall show, in fact, that this method enables us to find the integral in every finite interval in which the integral is continuous."

Again, the Encyclopédie des sciences mathématiques, in the article *Existence de l'intégrale générale. Détermination d'une intégrale particulière par ses valeurs initiales*, after going into some detail on the question of the length of the interval of convergence, says:‡

"On ne connaît encore aucun moyen de déterminer l'intervalle exact dans lequel la méthode de E. Picard converge. Suivant les cas, cet intervalle peut embrasser, comme dans la méthode de Cauchy-Lipschitz, tout l'intervalle de régularité de la solution, ou être au contraire plus petit que l'intervalle de convergence des séries de Taylor en $(x - x_0)$ qui représentent la solution, quand elle est holomorphe."

*Since this note was written, I have learned that Professor Wedderburn made essentially the same suggestion, in unpublished form, a number of years ago.

† Loc. cit., p. 73.

‡ Tome 2, volume 3, fascicule 1, pp. 14–15.

While the present remarks do not perhaps invalidate either of these statements, it does seem fair to say that they have a bearing on the comparison.

It is readily seen that the method outlined above can be extended to the case of a system of n differential equations in n unknown functions. If there are two equations, for example,

$$\frac{dy}{dx} = f(x, y, z), \qquad \frac{dz}{dx} = \varphi(x, y, z),$$

with right-hand members defined for $x_0 \leqq x \leqq x_0 + a$, $y_0 - b \leqq y \leqq y_0 + b$, $z_0 - c \leqq z \leqq z_0 + c$, it is possible to set

$$\begin{array}{ll} f(x, y, z) = f(x, y_0 + b, z), & y \geqq y_0 + b, \quad z_0 - c \leqq z \leqq z_0 + c; \\ f(x, y, z) = f(x, y, z_0 + c), & y_0 - b \leqq y \leqq y_0 + b, \quad z \geqq z_0 + c; \\ f(x, y, z) = f(x, y_0 + b, z_0 + c), & y \geqq y_0 + b, \quad z \geqq z_0 + c; \end{array}$$

and similarly in the other regions of the yz-plane, with a corresponding treatment for φ. More concisely, for any number of dimensions, the value of each function at any point outside its original domain of definition is to be the same as the value which it has at the nearest point of that domain.

The same method, though of course not the same formulas, can be used even if the original domain is not rectangular, provided that it has a moderately regular boundary, so that the functions can be extended across the boundary with the requisite degree of continuity.

THE UNIVERSITY OF MINNESOTA,
MINNEAPOLIS, MINN.

A FUNDAMENTAL SYSTEM OF COVARIANTS OF THE TERNARY CUBIC FORM.

By L. E. Dickson.

1. In many different mathematical investigations use is made of covariants of the ternary cubic form F. Less frequent use is made of the further concomitants involving line coördinates, and these will not be discussed here. The complete system of the 34 concomitants was obtained by symbolic methods by Clebsch and Gordan* and simpler by Gundelfinger.† They were exhibited in non-symbolic form by Cayley‡ for the canonical form $\Sigma a_i x_i^3 + 6l x_1 x_2 x_3$. Certain concomitants are obtained in the texts by Salmon, Elliott, and Weber, but no attempt is made to find a fundamental system.

The object of the present paper is to prove by an elementary method that a fundamental system of covariants of F is given by F, two invariants§ S and T, the Hessian H of F, the bordered Hessian determinant G, and the Jacobian J of F, H, G:

$$
(1)\qquad 6^3H = \begin{vmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{vmatrix}, \qquad 6^2G = \begin{vmatrix} F_{11} & F_{12} & F_{13} & H_1 \\ F_{21} & F_{22} & F_{23} & H_2 \\ F_{31} & F_{32} & F_{33} & H_3 \\ H_1 & H_2 & H_3 & 0 \end{vmatrix},
$$

$$
9J = \begin{vmatrix} F_1 & H_1 & G_1 \\ F_2 & H_2 & G_2 \\ F_3 & H_3 & G_3 \end{vmatrix},
$$

where F_{ij} denotes $\partial^2 F/\partial x_i \partial x_j$ and H_i denotes $\partial H/\partial x_i$. The method enables us to compute anew the expressions for S and T, and to deduce the syzygy (9) between them and the covariants.

2. The general ternary cubic form is

$$
F = a_0x^3 + 3b_0x^2y + 3c_0xy^2 + d_0y^3 + 3(a_1x^2 + 2b_1xy + c_1y^2)z + 3(a_2x + b_2y)z^2 + a_3z^3.
$$

The weight of any coefficient is its subscript; the various terms of any seminvariant (§ 3) are of equal weight.

* Math. Annalen, vol. 6, 1873, p. 436.

† *Ibid.*, vol. 4, 1871, p. 144.

‡ Amer. Jour. Math., vol. 4, 1881, p. 4; Coll. Math. Papers, XI, p. 342.

§ Given in full in Salmon's Higher Plane Curves, § 221; Cayley, Coll. Math. Papers, II, p. 325, where, in S, cf^2h is a misprint for cfh^2, while in the 8th line of the 4th column of T, h^2 is a misprint for k^2 in $chijk^2$, and in the 5th line of the 5th column, fil^4 is a misprint for fjl^4. In the third column of the Hessian, cij and fkl are misprints for cfj and gil.

Without altering x or y, replace z by $z + tx + my$. Then F is replaced by a like form with the coefficients

$$a_3' = a_3, \quad a_2' = a_2 + ta_3, \quad b_2' = b_2 + ma_3, \quad a_1' = a_1 + 2ta_2 + t^2a_3,$$
$$b_1' = b_1 + tb_2 + ma_2 + tma_3, \quad c_1' = c_1 + 2mb_2 + m^2a_3, \quad \cdots.$$

Invariants with respect to all such replacements are obtained by eliminating t and m:

$$a_3' = a_3, \quad a_1'a_3' - a_2'^2 = a_1a_3 - a_2^2, \quad b_1'a_3' - a_2'b_2' = b_1a_3 - a_2b_2, \quad \cdots.$$

Apart from a factor which is a power of a_3, these invariants are the values of $a_1', b_1', \cdots$ for $t = -a_2/a_3$, $m = -b_2/a_3$, which give $a_2' = b_2' = 0$.

Hence by the replacement of z by $z - xa_2/a_3 - yb_2/a_3$, F becomes

$$(2) \qquad a_3z^3 + 3zQ/a_3 + f/a_3^2,$$

such that the coefficients in

$$Q = Ax^2 + 2Bxy + Cy^2, \qquad f = ax^3 + 3bx^2y + 3cxy^2 + dy^3$$

are invariants of F with respect to all the transformations

$$(3) \qquad z' = z + tx + my,$$

and, conversely,* any polynomial invariant under these transformations is the quotient of a polynomial in $a_3, A, \cdots, d$ by a power of a_3. We find that

$$(4) \qquad \begin{aligned} A &= a_1a_3 - a_2^2, & a &= a_0a_3^2 - 3a_1a_2a_3 + 2a_2^3, \\ B &= b_1a_3 - a_2b_2, & b &= b_0a_3^2 - a_1b_2a_3 - 2b_1a_2a_3 + 2a_2^2b_2, \\ C &= c_1a_3 - b_2^2, & c &= c_0a_3^2 - 2b_1b_2a_3 - c_1a_2a_3 + 2a_2b_2^2, \\ & & d &= d_0a_3^2 - 3c_1b_2a_3 + 2b_2^3. \end{aligned}$$

3. By a seminvariant of F is meant a homogeneous isobaric polynomial in its coefficients which is invariant with respect to all transformations (3) as well as all linear transformations on x and y. Hence the seminvariants are functions of a_3 and the simultaneous invariants of Q and f.

A fundamental system of invariants of Q and f is known† (§ 7) to be formed by the following five invariants: the discriminant $\Delta = AC - B^2$ of Q, the discriminant

$$D = (ad - bc)^2 - 4(ac - b^2)(bd - c^2)$$

of f, the intermediate invariant‡

$$I = A(bd - c^2) - B(ad - bc) + C(ac - b^2)$$

* For binary forms, cf. Dickson, Algebraic Invariants, 1914, p. 47.

† Dickson, *ibid.*, p. 61; Salmon, Modern Higher Algebra, 4th ed., p. 187.

‡ That of Q and $Q' = A'x^2 + 2B'xy + C'y^2$ is $AC' - 2BB' + CA'$, given by the invariance of the discriminant of $Q + kQ'$.

between Q and the Hessian of f, the resultant R of Q and f, and the resultant M of two linear covariants, R and M being given in full by Salmon. They are connected by the syzygy

$$(5)\quad M^2 = -4\Delta^3 D^2 + D(R^2 + 12R\Delta I + 24\Delta^2 I^2) - 4RI^3 - 36\Delta I^4.$$

4. The expression obtained from $\Delta = AC - B^2$ by inserting the values (4) is seen to be divisible by a_3, the quotient being

$$(6)\qquad h = \begin{vmatrix} a_1 & b_1 & a_2 \\ b_1 & c_1 & b_2 \\ a_2 & b_2 & a_3 \end{vmatrix},$$

which is the leader (coefficient of z^3) of the Hessian of F. Similarly, we seek other combinations of Δ, D, I, R, M which are divisible by powers of a_3, in order to deduce a fundamental system of seminvariants. But to verify a relation between seminvariants, it is sufficient to prove it for the case in which

$$(7)\qquad a_2 = b_2 = a_1 = c_1 = 0,$$

since F can be transformed into a form satisfying (7) by means of transformations which leave all seminvariants unaltered; after obtaining (2), we have only to introduce the factors of Q as new variables x and y. For (7), we have

$$\Delta = -b_1^2 a_3^2, \qquad D = \{(a_0d_0 - b_0c_0)^2 - 4(a_0c_0 - b_0^2)(b_0d_0 - c_0^2)\}a_3^8,$$
$$I = -b_1(a_0d_0 - b_0c_0)a_3^5,\ R = -8a_0d_0b_1^3a_3^7,\ M = 8(a_0c_0^3 - b_0^3d_0)b_1^3a_3^{11},$$

while S, T and the leaders (§ 5) g, j of covariants G, J become

$$\begin{aligned} S &= a_0d_0a_3b_1 - b_1a_3b_0c_0 - b_1^4, \\ T &= D\,a_3^6 - (20a_0d_0 + 12b_0c_0)b_1^3a_3 - 8b_1^6, \\ g &= 8a_3^3b_0b_1^3c_0 + 9a_3^2b_1^6, \\ j &= -8a_3^5b_1^3(a_0c_0^3 - b_0^3d_0). \end{aligned}$$

By (6), $h = -a_3b_1^2$. Hence we have the relations*

$$(8)\qquad \begin{aligned} \Delta = a_3h,\ \ a_3^4S &= -I - \Delta^2, & a_3^6T &= D + 12\Delta I + 4R + 8\Delta^3, \\ a_3^4g &= -8\Delta I - R - 9\Delta^3, & a_3^6j &= -M. \end{aligned}$$

Since any seminvariant of F is the quotient of a polynomial in a_3, Δ, I, R, D, M by a power of a_3, it equals the quotient of a polynomial in a_3, h, S, g, T, j by a power of a_3. We may assume that the exponent of j is 0 or 1 in view of (5), or the equivalent syzygy obtained by inserting the values of Δ, I, R, D, M, and noting that the terms in a_3^9, a_3^{10}, a_3^{11} cancel:

* These were also verified for the case $a_2 = b_2 = b_0 = c_0 = 0$.

$$
\begin{aligned}
j^2 = & - 4a_3^5hS^4 - 4a_3^4(gS^3 + 2h^2S^2T) \\
& + a_3^3(108h^3S^3 - 4ghST - 4h^3T^2) \\
& + a_3^2(36gh^2S^2 + 108h^4ST + g^2T) \\
& - a_3(516h^5S^2 + 36g^2hS + 18gh^3T) \\
& + 108h^4gS - 27h^6T + 4g^3.
\end{aligned} \tag{9}
$$

The syzygy between the covariants is derived by replacing a_3, h, g, j by F, H, G, J.

To conclude that a fundamental system of seminvariants of F is given by a_3, h, g, j, S, T, it now suffices to verify that no polynomial in the last five, linear in j, is divisible by a_3. It suffices to show this when $a_1 = a_3 = b_1 = b_2 = c_0 = 0$, for which (§ 5)

$$
\begin{aligned}
h &= - a_2^2c_1, \quad g = a_2^6d_0^2, & j &= (- 2a_2^2d_0^3 - 27b_0c_1^4)a_2^7, \\
S &= a_0a_2c_1^2 + a_2^2b_0d_0, & T &= 4a_0a_2^3d_0^2 - 27a_2^2b_0^2c_1^2.
\end{aligned}
$$

No polynomial in h, g, S, T is identically zero, since the Jacobian of S and T with respect to a_0 and b_0 is not identically zero. Next, if $j\rho + \sigma = 0$, where ρ and σ are polynomials in h, g, S, T, we find by changing the signs of b_0 and d_0 that $- j\rho + \sigma = 0$, whence $\sigma \equiv \rho \equiv 0$. Since a covariant is uniquely determined by its leader, which is a seminvariant, the covariants mentioned in § 1 form a fundamental system.

5. To compute the leaders g and j of our covariants G and J, we need certain coefficients of the Hessian:

$$H = Ex^2z + Fxyz + Py^2z + Qxz^2 + Kyz^2 + Lz^3 + \cdots.$$

Then the coefficient of z^6 in G is

$$
\begin{gathered}
g = Q^2\gamma + 2QK\delta + K^2\epsilon + 6QL\kappa + 6LK\lambda + 9L^2\mu, \\
\gamma = b_2^2 - a_3c_1, \qquad \delta = a_3b_1 - a_2b_2, \qquad \epsilon = a_2^2 - a_1a_3, \qquad \kappa = a_2c_1 - b_1b_2, \\
\lambda = a_1b_2 - a_2b_1, \qquad \mu = b_1^2 - a_1c_1.
\end{gathered}
$$

The coefficients of xz^5 and yz^5 in G are respectively

$$
\begin{aligned}
r = & \; Q^2(2b_1b_2 - a_2c_1 - a_3c_0) + 4EQ\gamma + 2QK(a_3b_0 - a_1b_2) + (2QF + 4EK)\delta \\
& + K^2(a_1a_2 - a_0a_3) + 2FK\epsilon + 6QL(a_2c_0 - b_0b_2 + \mu) + (4Q^2 + 12EL)\kappa \\
& + 6LK(a_0b_2 - a_2b_0) + (6LF + 4QK)\lambda + 9L^2(2b_0b_1 - a_0c_1 - a_1c_0), \\
w = & \; Q^2(b_2c_1 - a_3d_0) + 2QF\gamma + 2QK(a_3c_0 - a_2c_1) + (2FK + 4QP)\delta \\
& + K^2(2a_2b_1 - a_1b_2 - a_3b_0) + 4PK\epsilon + 6QL(a_2d_0 - b_2c_0) \\
& + (4QK + 6FL)\kappa + 6LK(b_0b_2 - a_2c_0 + \mu) \\
& + (12LP + 4K^2)\lambda + 9L^2(2b_1c_0 - a_1d_0 - b_0c_1).
\end{aligned}
$$

Then

$$
3j = \begin{vmatrix} a_2 & Q & r \\ b_2 & K & w \\ a_3 & 3L & 6g \end{vmatrix}.
$$

6. If we seek all the concomitants of F, viz., the covariants of F and a linear form L, let the transformation which reduces F to (2) replace L by $kz + l$, where l is linear in x, y. Hence we need the invariants of l, Q, f, viz., the covariants (or seminvariants) of Q and f. Although the latter are known (§ 7) and various concomitants of F can be readily deduced, the work of deriving a fundamental system and especially the proof that it is complete would seem prohibitive by this method.

7. If we seek a fundamental system of seminvariants of the binary quadratic form Q and cubic form f, given in § 2, we begin by removing the second term of f by replacing x by $x - yb/a$. Then f and Q become

$$ax^3 + \frac{3}{a}A_{22}xy^2 + \frac{1}{a^2}A_{33}y^3, \qquad Ax^2 - \frac{2}{a}B_{13}xy + \frac{1}{a^2}(aB_{11} - AA_{22})y^2,$$

where

$$A_{22} = ac - b^2, \quad A_{33} = a^2d - 3abc + 2b^3, \quad B_{13} = Ab - Ba, \quad B_{11} = Ac - 2Bb + Ca.$$

Hence every seminvariant is the quotient of a polynomial in $A_{13} = a$, A_{22}, A_{33}, $B_{02} = A$, B_{11}, B_{13} by a power of A_{13}. Among these quotients is the discriminant $A_{40} = D$ of f given by the syzygy

$$A_{13}^2A_{40} - 4A_{22}^3 - A_{33}^2 = 0.$$

Other quotients B_{22}, B_{31}, $B_{20} = I$, C_{11}, $C_{00} = \Delta$, $C_{31} = L_4 + IL_1$, $D_{20} = R + 8\Delta I$, $D_{40} = M$ are defined in turn by Hammond's* syzygies (2), (3), (4), (8), (9), (13), (23), (27) between our 15 seminvariants. He listed 35 further syzygies deducible from these nine. These 44 syzygies might be used to simplify any polynomial in the 15 seminvariants in an attempt to prove that the simplified polynomial, regarded as a function of $A, \cdots, d$, does not have the factor $A_{13} = a$, unless the initial polynomial has the explicit factor A_{13}, and hence to prove that the 15 forms give a fundamental system. To indicate only one step in this rather prohibitive work, we first eliminate the products of D_{40} by A_{13}, A_{22}, A_{33}, B_{02}, B_{11}, B_{13}, B_{22}, B_{31}, C_{11}, C_{31}, D_{40} by means of syzygies (27), (30), (32), (35), (37), (38), (39), (41), (42), (43), (44); then D_{40} occurs only with invariants A_{40}, B_{20}, C_{00}, D_{20}. Since D_{40} is the only one of these invariants which is skew (of odd weight), it cannot occur in the polynomial.

Such a proof, if completed, would also yield a complete set of syzygies. The proof that the 15 covariants form a fundamental system is however much simpler by the symbolic theory.†

* Amer. Jour. Math., vol. 8, 1886, p. 138. His notation for a covariant has been retained here for its seminvariant leader. In verifying a syzygy between the latter, we may take $b = 0$.

† Clebsch, Binaren Algebraischen Formen, 1872, p. 209; Glenn, The Theory of Invariants, 1915, p. 146.

THE MODULAR THEORY OF POLYADIC NUMBERS.

By Albert A. Bennett.

1. **Introduction.** While the study of numbers usually involves the frequent representation of integers in one or another notational system, it is not ordinarily the properties of the method of representation that are studied but rather the intrinsic features of the numbers themselves. It is nevertheless true for example that the time spent in teaching elementary arithmetic for commercial purposes involves not a few hours devoted to the mere technique of manipulating Arabic decimal symbols. The distinction drawn in elementary text books between decimals and fractions is based on the existence of two methods of representing rational numbers and of two partially distinguishable types of problems, each being dealt with more easily in one notation than in the other. Were the classical Roman notation the only one employed, several chapters in the grade school text books on arithmetic would be appearing in very different form. It will be instructive to keep this fact in mind in reading this article, for this discussion will involve properties of numbers in part dependent on a representation similar to the decimal or, more properly speaking, decadic notation. On this account a few words will first be said concerning the decadic notation.

A number in the decadic notation, using Arabic symbols for the digits, is denoted by a sequence of digits, the sequence being not necessarily terminating. A non-integral rational number usually requires a decimal point in its decadic representation. Thus $\frac{1}{2} = .5$, $1/100 = .01$, $\frac{1}{3} = .333 \cdots$, while an integer does not require a decimal point, the figures to the right of the decimal point when written being all zeros. Numerical symbols employing the decimal point fall into two logical classes. In one class the decimal "terminates," that is, after a finite number of digits, the figures form an unbroken sequence of zeros, which zeros need not be expressed. In the other the figures do not "terminate" and no matter how far out one proceeds to the right of the decimal point digits other than zero may be found still further out in each of the expressions of the second class. A number is representable as a terminating decimal if, and only if, it is expressible as P/Q where P and Q are relatively prime and where Q is a factor of some power of 10, the base of the decadic notation. On the other hand, every positive number is representable as a non-

terminating decimal by using the fact that

$$1. = .9999 \cdots.$$

Thus $1/2 = .49999 \cdots$, $1/100 = .009999 \cdots$.

Despite one's familiarity in elementary instruction with the idea of a non-terminating decimal, it is ordinarily assumed that no representation which is non-terminating to the left is to be employed. This is however a matter rather of custom and convenience than of logical requirement although a consideration of relative magnitude makes it generally desirable. For example, the notation

$$\cdots,333,334. = x$$

can represent but a single number, which may be readily identified in a few simple steps. If the product $3x$ be formed, it is found to be identically of the form, $\cdots,000,002$, so that x represents, if anything, the number $2/3$.

We shall define a *decadic integer* as a decadic symbol whether or not terminating to the left, but not requiring a decimal point. Thus $.6666\cdots$ is not a decadic integer while $\cdots,333,334$ is a decadic integer although both express the fraction $2/3$. It may be readily proved that every positive rational number P/Q when Q is not a factor of any power of the base, 10, may be written as a decadic integer. It is further to be noted that the negative of a decadic integer is also a decadic integer. For example,

$$-1 = \cdots,999,999.$$

Indeed, decadic integers may be added, subtracted or multiplied with decadic integers for results. Thus decadic integers constitute what is called a domain of integrity. Unlike the case of decimals, the representation of a number as a decadic integer, when at all possible, is unique.

2. Polyadic numbers. A generalization from the decadic representation to a *b-adic representation*, where b is an arbitrary integer greater than unity, involves no difficulties. While a number in decadic representation is also capable of representation in a b-adic system, a decadic integer is not necessarily a b-adic integer. For example, the decadic integer $\cdots,333,334$ cannot be a 3-adic integer since the denominator of $2/3$ is not prime to the base 3. A number may be frequently expanded as an integer simultaneously with respect to n distinct bases and so be represented as a b_1-adic, b_2-adic, $\cdots$, b_n-adic integer. If the bases, b_1, b_2, $\cdots$, b_n, be known, any one representation of course determines the number itself and therefore the other representations also.

A set of independent expressions $(a_1, a_2, \cdots, a_n)$, where a_i is a b_i-adic integer, $i = 1, 2, \cdots, n$, may be studied as in the theory of complex

numbers where in particular it may happen that the n symbols denote the same abstract number. By the sum of two such symbols $(a_1, a_2, \cdots, a_n)$ and $(a_1', a_2', \cdots, a_n')$ will be meant the symbol $(a_1 + a_1', a_2 + a_2', \cdots, a_n + a_n')$ and by their product will be meant $(a_1 \times a_1', a_2 \times a_2', \cdots, a_n \times a_n')$. Such a symbol may be called a *polyadic number* and in particular where each element is a respective b-adic integer, the polyadic number is called a *polyadic integer*. In the same manner, b-adic integers to a given set of bases, $b_1, b_2, \cdots, b_n$, constitute a domain of integrity.

When b is a composite number equal to, say, $p_1^{m_1}p_2^{m_2}\cdots p_k^{m_k}$, where the p's are distinct primes, the study of the single b-adic numbers is much enriched by considering their polyadic representations with respect to the k bases, $p_1, p_2, \cdots, p_k$. It is to be noticed that any number which is expressible as a b-adic integer will also be integral in this polyadic representation and the converse is also true. The theory of b-adic numbers may therefore be confined in the first instance to cases where b is a prime, the composite cases being included in the theory of polyadic numbers where each base is a prime.

If $\pi_1, \pi_2, \pi_3, \cdots$ be the successive primes 2, 3, 5, $\cdots$, as occurring in their order of magnitude, the polyadic numbers with an infinite number of bases, $\pi_1, \pi_2, \pi_3, \cdots$, may be considered. Any other system will be a *section* of the system so obtained, being the result of omitting some of the bases from this system. This "complete" polyadic system of integers consists therefore of numbers which may be represented by an array

$$\begin{array}{c}
\cdots a_{m1} \cdots a_{31}a_{21}a_{11}a_{01} \\
\cdots a_{m2} \cdots a_{32}a_{22}a_{12}a_{02} \\
\cdots a_{m3} \cdots a_{33}a_{23}a_{13}a_{03} \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdots a_{mn} \cdots a_{3n}a_{2n}a_{1n}a_{0n} \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot
\end{array}$$

where the nth row denotes the π_n-adic number indicated by the notation, $\cdots + a_{mn}\pi_n^m + \cdots + a_{3n}\pi_n^3 + a_{2n}\pi_n^2 + a_{1n}\pi_n + a_{0n}$.

Not only may sections of the array be considered which correspond to the suppression of certain entire rows, but a more general type of section is of interest where left-hand portions of some rows are suppressed. This may be illustrated in the case of a single p-adic number where p is any prime number.

3. **Modular b-adic numbers.** Let q be a positive integer and b be any base, and consider the section of b-adic numbers including through the coefficients of b^{q-1} only, all consideration of the coefficient of b^q and

higher powers being omitted. In other words, consider the modular theory of b-adic integers, modulo b^q. Addition, subtraction and multiplication may be effected by the usual rules so that the modular b-adic numbers (*mod.* b^q) of themselves constitute a domain of integrity. The theory of b-adic numbers to the modulus b^q, where $b = p_1^{m_1}p_2^{m_2}\cdots p_k^{m_k}$, is included in the theory of polyadic numbers, to the bases, $p_1, p_2, \cdots, p_k$, and to the respective moduli, $p_1^{qm_1}, p_2^{qm_2}, \cdots, p_k^{qm_k}$, which is the study of a section of the general double array of coefficients.

More generally, it is always possible to find a single positive integer whose expansion with respect to n distinct prime numbers, $p_1, p_2, \cdots, p_n$ (n, finite), shall coincide with a given section including only these bases taken *modd.* $p_1^{m_1}, p_2^{m_2}, \cdots, p_n^{m_n}$. Two such numbers will differ in fact by an integral multiple of the product, $p_1^{m_1} \times p_2^{m_2} \times \cdots \times p_n^{m_n}$. This situation no longer persists when n is allowed to become infinite. However, the array may still be treated as a "number" in the sense of a complex number or for n infinite it may be thought of as a sort of fictitious limiting number. Similar remarks apply when one of the exponents, m_i, is allowed to increase indefinitely.

4. Division among b-adic numbers. In a b-adic integer, expressed by the symbol, $\cdots a_m \cdots a_3a_2a_1a_0$, or more explicitly in the form, $\cdots + a_mp^m + \cdots + a_3p^3 + a_2p^2 + a_1p + a_0$, the term a_0 is called the *principal* term of the number. A b-adic integer is said to be *singular* or *nonsingular* according as its principal term does or does not vanish. The words "singular" and "nonsingular" are applied directly to b-adic integers only when b is a prime. For composite bases, b, the representation as a polyadic number to bases which are the distinct prime factors of b is chosen. A polyadic number is singular if any one of its principal terms is zero, and it is nonsingular if, and only if, none of its principal terms vanishes.

Division giving a unique quotient is possible by polyadic numbers with prime bases whether or not in a modular domain if and only if these be nonsingular. In the complete domain, that is, where no modular reductions are made, division by a singular number, no row of which is entirely zero, is possible by the introduction of terms to the right of the "decimal" point, that is, by going outside of the integral domain. In particular, nonsingular polyadic numbers have nonsingular reciprocals. The nonsingular numbers do not form, however, a domain of integrity, since the sum of two nonsingular numbers may be singular.

5. Elementary units. The modular theory of polyadic numbers presents little of interest not found in the simple case of a single p-adic number. Some of the few particular points worth mentioning may be given here. An *elementary unit* is defined as a number whose polyadic

representation contains the figure one as a principal term in one row, all other figures of this row and of other rows being zero. Let the modular polyadic numbers be taken *modulis* $p_1^{m_1}$, $p_2^{m_2}$, $\cdots$, $p_k^{m_k}$, and let it be desired to identify the elementary unit which has 1 for the principal term of the row for p_i. Such a number, e_i, has the property that $e_i = 0$, *modd.* $p_1^{m_1}$, $p_2^{m_2}$, $\cdots$, $p_{i-1}^{m_{i-1}}$, $p_{i+1}^{m_{i+1}}$, $\cdots$, $p_n^{m_n}$, while $e_i = 1$, *mod.* $p_i^{m_i}$. As is well known Euclid's algorism for the highest common factor may be applied to a pair of relatively prime numbers P and Q, so as to determine two integers M and N of opposite sign such that

$$MP + NQ = 1,$$

where furthermore M is numerically less than Q and N numerically less than P, so that MP is equal to 1 *modulo* Q and is equal to 0 *modulo* P. By taking P as the product, $p_1^{m_1} \times p_2^{m_2} \times \cdots \times p_{i-1}^{m_{i-1}} \times p_{i+1}^{m_{i+1}} \times \cdots \times p_n^{m_n}$, and Q as $p_i^{m_i}$, an $MP = e_i$ is obtained. The successive elementary units e_1, e_2, $\cdots$, e_n having been determined, we have $e_1 + e_2 + \cdots + e_n = 1$ (*mod.* b), where $b = p_1^{m_1} \times p_2^{m_2} \times \cdots \times p_n^{m_n}$. The theory of numbers to a composite modulus is fairly illustrated in the case of the modulus $12 = 2^2 \times 3$. The twelve numbers of the set may be represented as follows,

$$1 = \begin{pmatrix} 0 & 1 \\ & 1 \end{pmatrix}, \quad 2 = \begin{pmatrix} 1 & 0 \\ & 2 \end{pmatrix}, \quad 3 = \begin{pmatrix} 1 & 1 \\ & 0 \end{pmatrix}, \quad 4 = \begin{pmatrix} 0 & 0 \\ & 1 \end{pmatrix},$$

$$5 = \begin{pmatrix} 0 & 1 \\ & 2 \end{pmatrix}, \quad 6 = \begin{pmatrix} 1 & 0 \\ & 0 \end{pmatrix}, \quad 7 = \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}, \quad 8 = \begin{pmatrix} 0 & 0 \\ & 2 \end{pmatrix},$$

$$9 = \begin{pmatrix} 0 & 1 \\ & 0 \end{pmatrix}, \quad 10 = \begin{pmatrix} 1 & 0 \\ & 1 \end{pmatrix}, \quad 11 = \begin{pmatrix} 1 & 1 \\ & 2 \end{pmatrix}, \quad 12 = 0 = \begin{pmatrix} 0 & 0 \\ & 0 \end{pmatrix}.$$

The elementary units are 9, $= \begin{pmatrix} 0 & 1 \\ & 0 \end{pmatrix}$, and 4, $= \begin{pmatrix} 0 & 0 \\ & 1 \end{pmatrix}$. The upper row is always added 2-adically and the lower 3-adically. The singular numbers are 0, 2, 3, 4, 6, 8, 9, 10. The four nonsingular numbers, 1, 5, 7, 11, are each self-reciprocal. The nonsingular numbers always form a group under multiplication so that a multiplication table of the nonsingular numbers is always of interest—it includes, of course, in particular the reciprocal of each nonsingular number in every case. For the case of 12 as above we have as a multiplication table for the nonsingular numbers, the following self-explanatory tabulation

	1	5	7	11
1	1	5	7	11
5	5	1	11	7
7	7	11	1	5
11	11	7	5	1

6. Equivalence and singular classes. Two polyadic numbers are said to be *equivalent* if one is obtained from the other by multiplication by a nonsingular number. In particular, all nonsingular numbers in a given system are equivalent. In general, however, not all singular numbers are equivalent. The test for equivalence is obvious in the typical polyadic expansion. Two polyadic expansions are equivalent when corresponding rows have the same number of consecutive zeros counting from the right. Thus in the above

$$\begin{pmatrix} 1\,0 \\ \;\;2 \end{pmatrix}, \begin{pmatrix} 1\,0 \\ \;\;1 \end{pmatrix}, \text{ are equivalent,}$$

$$\begin{pmatrix} 1\,1 \\ \;\;0 \end{pmatrix}, \begin{pmatrix} 0\,1 \\ \;\;0 \end{pmatrix}, \text{ are equivalent,}$$

$$\begin{pmatrix} 0\,0 \\ \;\;1 \end{pmatrix}, \begin{pmatrix} 0\,0 \\ \;\;2 \end{pmatrix}, \text{ are equivalent,}$$

$$\begin{pmatrix} 0\,1 \\ \;\;1 \end{pmatrix}, \begin{pmatrix} 0\,1 \\ \;\;2 \end{pmatrix}, \begin{pmatrix} 1\,1 \\ \;\;1 \end{pmatrix}, \begin{pmatrix} 1\,1 \\ \;\;2 \end{pmatrix}, \text{ are equivalent,}$$

$$\begin{pmatrix} 0\,0 \\ \;\;0 \end{pmatrix} \text{ is only self-equivalent,}$$

and the five sets above are mutually non-quivalent.

For any singular polyadic number, a, other than zero there is a corresponding *singular class* $SC(a)$ within which division by a is unique; this singular class consists of all numbers of the total class which have at least the initial zeros of a. Thus if in a in the ith row the first j figures counted from the right are zero then each polyadic number in $SC(a)$ will have zeros for the first j figures from the right in the ith row. The analogue of the singular class for a nonsingular number is the complete class. In the example above we have

$$\begin{aligned} SC\,(0) &= 0, \\ SC\,(2) &= 0, 2, 4, 6, 8, 10, \\ SC\,(3) &= 0, 3, 6, 9, \\ SC\,(4) &= 0, 4, 8, \\ SC\,(6) &= 0, 6, \\ SC\,(8) &= SC\,(4), \\ SC\,(9) &= SC\,(3), \\ SC\,(10) &= SC\,(2). \end{aligned}$$

A singular class is always a domain of integrity and is sometimes a field, that is, division within a singular class is sometimes possible by every number other than zero. The numbers common to two singular classes always constitute a singular class. A singular class whose only singular

subclasses are itself and zero is called a *primitive singular subclass.* A singular subclass other than 0 is a field if and only if it is primitive. A primitive subclass consists of numbers whose polyadic expansions consist, except for a single common fixed element, wholly of zeros. There are therefore as many primitive singular subclasses as there are rows in the expansion. A primitive singular unit is defined as $p_i^{m_i-1}e_i$ where e_i is the elementary unit for the modulus, $p_i^{m_i}$. A primitive singular unit generates a primitive singular subclass. The primitive singular units for the modulus twelve are

$$\begin{pmatrix} 1 & 0 \\ & 0 \end{pmatrix} = 6 \text{ and } \begin{pmatrix} 0 & 0 \\ & 1 \end{pmatrix} = 4,$$

where 4 is also an elementary unit.

7. **Sets of p-adic numbers with a common base.** We shall now turn to the case of sets of p-adic numbers proper with a single common base rather than polyadic numbers with several distinct bases.

Let $(a_0, a_1, a_2, \cdots, a_n)$ be a set of $(n + 1)$ p-adic integers, with the common base p, a prime. This set will be called *singular* if and only if *each* of the numbers $a_0, a_1, \cdots, a_n$ is singular. A single set will be said to be of *nullity* s, if there is at least one of the numbers in which the coefficient of p^s does not vanish while the coefficients of p^i, $i = 0, 1, 2, \cdots, s - 1$, vanish for each of the $n + 1$ numbers of the set. A nonsingular set may be called the *point coördinates* of a point in a *p-adic projective system* provided that two sets are regarded as corresponding to the same point if and only if these sets may be obtained one from the other by multiplication by a nonsingular p-adic number as a factor.

Two points are said to be *neighboring* if a singular set other than zero is linearly dependent on the coördinates of the two respective sets of coördinates. Two points are said to be in a *neighborhood of the sth order,* when a singular set of nullity s is linearly dependent upon their coördinates.

A set of points is *linearly dependent* if the null set, zero, may be expressed as a linear combination of them with nonsingular coefficients. The set is *linearly semi-dependent of order s* when a singular set of nullity, s, may be represented as a linear combination of them with nonsingular coefficients but no singular set of nullity greater than s is so expressible. When no singular set is expressible as a linear combination of the given set with nonsingular coefficients—what may be thought of as semi-dependence of order zero—the given system is *linearly independent.* Thus two semi-dependent points are neighboring.

With these concepts and definitions one may study a modular "geometry" with a composite modulus of the form p^m. This is to be

distinguished from the Galois field of p^m where the modulus is merely p. The Galois field is obtained by introducing algebraic irrationalities in the field of p itself. The domain here treated, of p^m, is not a field and in consequence does not yield so natural a form of "geometry." In fact, the limiting type of geometry for s kept fast and greater than unity, while p is allowed to increase indefinitely, is not the usual real geometry but a non-Archimedean theory with actual constant "infinitesimals." For composite numbers not merely powers of primes, the geometrical study is better carried out by considering separately the distinct relatively prime factors each of the form p^s.

A discussion of some related ideas will be found in Fraenkel, Teiler der Null und Zerlegung von Ringen, Journ. f. d. r. u. ang. Math. (Crelle), (145) 1915, (139–176). The notion of p-adic integers is due to K. Hensel. References will be found in the above paper.

BALTIMORE, MD.,
September, 1920.

SOME ALGEBRAIC ANALOGIES IN MATRIC THEORY.

By Albert A. Bennett.

An obvious analogy exists between the theory of matrices and the theory of algebraic numbers. The analogy is in some respects superficial, but it is suggestive and extends further than is usually pointed out. A conspicuous cause of difference in the two theories is that while multiplication among algebraic numbers is always commutative, this is not the case among square matrices of a given order. As a result, a matric equation with scalar coefficients when satisfied by a given matrix is satisfied also by all transforms of this matrix through nonsingular matrices. The number of nonsingular distinct roots cannot usually be finite.

In the following discussion the matrices considered will be assumed without further mention to be square matrices and all of the same order. Such theorems concerning matrices as are found in Bôcher's " Introduction to Higher Algebra " will be assumed without discussion. The term " conjugate " as applied to a matrix will not be used in the current sense of the transposed matrix, obtained by turning the given matrix over about its main diagonal and thus interchanging rows and columns. On the contrary by " conjugate " will be meant the algebraic analogue of the term as used in the theory of algebraic numbers and given for matrices explicitly in detail by H. Taber.* The term " scalar " will be applied to a matrix having zeros except in the main diagonal and having the elements in the main diagonal equal. The " latent roots " of a matrix, or roots of the characteristic equation of a matrix will be called the *characteristic numbers* of the matrix.

Some Theorems Concerning Matrices Which Have Immediate Algebraic Analogues.

We shall list below a set of twenty-eight propositions concerning matrices, each of which may be translated at once into its counterpart in the theory of algebraic numbers. To do this it is merely necessary to substitute as follows:

For "identical matrix," I, substitute "unity" (1).

For "null matrix," 0, substitute "zero," 0.

For "scalar," substitute "rational number."

* H. Taber, On certain identities in the theory of matrices. Amer. Journ. Math., vol. 13 (1891), pp. 159–172.

For "number," substitute "integer."

For "matrix," substitute "algebraic number."

For "matrix with distinct non-vanishing characteristic numbers," substitute "Galoisian algebraic numbers."

For "characteristic function," substitute "defining function."

For "determinant," substitute "norm."

1. The identical matrix, I, and the null matrix, 0, are scalars.

2. Addition, subtraction, multiplication and division according to the usual rules of algebra may be performed among scalars.

3. The matric equation $ax = bI$ where a and b are numbers, and a is not zero, has a unique scalar as a solution, and each scalar is the root of such an equation.

4. If α is a non-scalar matrix there exists a polynomial

$$f(x) = x^n - s_1x^{n-1} + \cdots + (-1)^n s_n,$$

with scalar coefficients, of which α is a root.

5. There is a minimum degree (> 1) for such a function and there is but one function of this minimum degree.

6. Certain matrices are distinguished by many simple properties and are worthy of special study. For the present, only matrices with distinct nonvanishing characteristic numbers will be discussed, although some of the relations mentioned apply to all matrices.

7. The minimum degree n of the $f(x)$ for a matrix, α, of distinct non-vanishing characteristic numbers is called the *order* of α, and $f(x)$, *its characteristic function.*

8. The characteristic function, $f(x)$, of a matrix α of distinct non-vanishing characteristic numbers has a set of n distinct roots, $\alpha, \alpha_1, \alpha_2, \cdots, \alpha_{n-1}$, where $\alpha_1, \alpha_2, \cdots, \alpha_{n-1}$ are called the *conjugates* of α, and these satisfy the following conditions:

(i) Each conjugate, α_i, may be expressed as a polynomial in α with scalar coefficients.

(ii) Each conjugate, α_i, is a matrix of the same order, n, and with the same characteristic function, $f(x)$, as α.

(iii) The elementary symmetric functions of the set $(\alpha, \alpha_1, \cdots, \alpha_{n-1})$ are (except for sign) the n scalar coefficients $s_1, s_2, \cdots, s_n$ of the characteristic function,

$$f(x) = x^n - s_1x^{n-1} + \cdots + (-1)^n s_n.$$

9. The coefficient s_1 is called the *trace* of α, and the coefficient s_n, the *determinant* of α.

10. The function $f(x)$ may be viewed as the determinant of $(x - \alpha)$.

11. For α, a matrix with distinct nonvanishing characteristic numbers,

it is possible to select in many ways a *basis* of n matrices $\beta_1, \beta_2, \cdots, \beta_n$, linearly independent polynomials in α, with scalar coefficients, such that the totality of linear combinations with scalar coefficients, of the matrices of the basis, include all rational functions of α, where the indicated division has a meaning.

12. Two possible choices of a basis are

$$(1, \alpha, \alpha^2, \cdots, \alpha^{n-1}) \qquad \text{and} \qquad (\alpha, \alpha_1, \alpha_2, \cdots, \alpha_{n-1}).$$

13. In particular, for α, a matrix with distinct nonvanishing characteristic numbers, every rational function of α, where the indicated division results in a finite matrix and where the coefficients are scalars, is expressible as a polynomial in α of degrees less than n, with scalar coefficients.

14. The totality of such rational functions of α may be called the *domain* of α. Multiplication within the domain is commutative.

15. For any n matrices, $\gamma_1, \gamma_2, \cdots, \gamma_n$, of the domain of a matrix α of distinct nonvanishing characteristic numbers, the discriminant of $(\gamma_1, \gamma_2, \cdots, \gamma_n)$ is defined as the determinant of scalars,

$$\begin{vmatrix} S(\gamma_1\gamma_1) & S(\gamma_1\gamma_2) & \cdots & S(\gamma_1\gamma_n) \\ S(\gamma_2\gamma_1) & S(\gamma_2\gamma_2) & \cdots & S(\gamma_2\gamma_n) \\ \cdot & \cdot & \cdot & \cdot \\ S(\gamma_n\gamma_1) & S(\gamma_n\gamma_2) & \cdots & S(\gamma_n\gamma_n) \end{vmatrix},$$

where $S(\xi)$ is the *trace* of ξ. The discriminant is denoted by the symbol,

$$\Delta(\gamma_1, \gamma_2, \cdots, \gamma_n).$$

16. If $\gamma_i = \Sigma_j r_{ij}\beta_j$, where r_{ij} is scalar, then $\gamma_i\gamma_k = \Sigma_j r_{ij}\beta_j \Sigma_l r_{kl}\beta_l = \Sigma_{jl}(r_{ij}r_{kl})(\beta_j\beta_l)$. But $S(r\delta) = rS(\delta)$, where r is scalar, and $S(\delta_1 + \delta_2) = S(\delta_1) + S(\delta_2)$ for $\delta, \delta_1, \delta_2$, any matrices of the domain.

$$\therefore S(\gamma_i\gamma_k) = \Sigma_{jl}(r_{ij}r_{kl})S(\beta_j\beta_l).$$

By reference to the rule for multiplication of determinants, we have

$$\Delta(\gamma_1, \gamma_2, \cdots, \gamma_n) = [\text{Det }(r_{ij})]^2\Delta(\beta_1, \beta_2, \cdots, \beta_n).$$

17. The discriminant of the basis $(\alpha, \alpha_1, \cdots, \alpha_{n-1})$ is not zero. It is expressible as

$$\begin{vmatrix} \alpha & \alpha_1 & \alpha_2 & \cdots & \alpha_{n-2} & \alpha_{n-1} \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{n-1} & \alpha \\ \alpha_2 & \alpha_3 & \alpha_4 & \cdots & \alpha & \alpha_1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha_{n-1} & \alpha & \alpha_1 & \cdots & \alpha_{n-3} & \alpha_{n-2} \end{vmatrix}^2.$$

18. Hence the discriminant of every basis of the domain is different from zero.

19. It is possible to find a matrix α, of order n, no restriction as to the characteristic numbers being imposed, such that the equation $x^2 = \alpha$ is not satisfied by any proper matrix of order n.

20. In order to render certain general matric theorems as to the existence of a matric equation universally valid, it is sometimes necessary to introduce an improper root, which may be viewed as the limit of a finite matrix, as a convenient parameter approaches infinity.

21. The product of the $n - 1$ conjugates of α is a matrix of the domain of α, called the *adjoint* of α, $A(\alpha)$.

22. The determinant of the adjoint is

$$AA_1A_2\cdots A_{n-1} = A(\alpha)A(\alpha_1)\cdots A(\alpha_{n-1}) = (\alpha\alpha_1\cdots\alpha_{n-1})^{n-1},$$

which is the $(n - 1)$st power of the determinant of α.

23. The adjoint of the adjoint of α is in the same manner equal to α times the $(n - 2)$nd power of the determinant of α.

24. The sum of the $(n - 1)$ conjugates of α is a matrix of the domain of α called the adjoint-trace of α, $T(\alpha)$.

25. The trace of the adjoint-trace of α is $(n - 1)$ times the trace of α.

26. The adjoint-trace of the adjoint-trace of α is α plus $(n - 2)$ times the trace of α.

27. If γ is any matrix of the domain of α, the adjoint of $l - \gamma$, where l is scalar, is a polynomial in l of degree $n - 1$ with the coefficients in the domain.

28. If γ is any matrix of the domain of α, the determinant of $l - \gamma$, where l is scalar, is a polynomial in l of degree n with scalar coefficients.

A Discussion of Improper or Limit Matrices.

Any square matrix may be obtained as the limit of a matrix with distinct nonvanishing characteristic numbers, and theorems for a general matrix may sometimes be obtained by passage to a limit from this restricted but important case. It is needless to insist that care must be exercised. The well-known theorem that all matrices commutative with respect to multiplication with a given matrix of distinct nonvanishing characteristic numbers are rational integral functions of the given matrix has sometimes been stated for the general matrix. The theorem is, however, false, as is seen by reference to the matrix $\binom{1\ 0}{0\ 1}$. Some of the elements of a matrix β, which is obtained from a given matrix α of distinct nonvanishing characteristic numbers, may become infinite as two of the characteristic numbers of α approach equality, or one approaches zero. The limit may lead, therefore, not to a proper matrix but to an improper or limit matrix containing infinite elements.

An explicit mention of a similar limiting case is found in the classical memoir by Frobenius.* On pages 43 and 44 is found the following:

I. Every substitution, U (of determinant, $+1$), which transforms into itself a symmetric form, S, of nonvanishing determinant and for which the determinant of $E + U$ vanishes, may be expressed in the form

$$U = \lim (h = 0), \quad (S + T_h)^{-1}(S - T_h),$$

where T_h is an alternating form whose coefficients are rational functions of h.

II. Every substitution, U, which transforms into itself an alternating form, T, of nonvanishing determinant, and for which the determinant of $E - U$ vanishes, may be expressed in the form

$$U = \lim (h = 0), \quad (S_h + T)^{-1}(S_h - T),$$

where S_h is a symmetric form whose coefficients are rational functions of h.

Another occasion for the use of improper matrices is in the extraction of square roots of matrices. While for e, different from zero, $\begin{pmatrix} e^2 & 2 \\ 0 & e^2 \end{pmatrix}$ has the square root $\begin{pmatrix} e & 1/e \\ 0 & e \end{pmatrix}$, yet for $e = 0$, there is no proper matrix obtained as a square root but only an improper matrix, as a limit.

The relations between the trace, adjoint-trace, determinant, and adjoint may be so expressed as to be valid for all square matrices without restriction as to characteristic numbers. Thus there are certain relations which in terms of the conjugates of a matrix become obvious but which are capable of proof without reference to conjugates. Many of the arguments which have resulted in the successive historical extensions of the number system and in the introduction as valid numbers of negatives, fractions, irrationals, imaginaries, may be urged for the acceptance of limit matrices, at least when these are required to render general the notion of conjugates.

The matrix, α, taken as

$$\begin{pmatrix} e & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

for $e \neq 1, \neq 2$, has two proper conjugates in the sense used above, which may be taken as

$$\begin{bmatrix} 2 & -\dfrac{2-e}{1-e} & 0 \\ 0 & e & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & \dfrac{1}{1-e} & 0 \\ 0 & 2 & 0 \\ 0 & 0 & e \end{bmatrix},$$

* Frobenius, Über lineare Substitutionen und bilineare Formen. Jour. f. d. reine und ang. Math., vol. 84 (1878), pp. 1–63.

which become improper as e approaches unity. For $e = 1$, there is no set of proper conjugates. It will not be sufficient to denote both lim $(e = 1)$, $-(2 - e)/(1 - e)$, and lim $(e = 1)$, $1/(1 - e)$ by the mere sign ∞. The algebraic relations between these quantities must be retained also in the limit. Despite these difficulties, symbols α_1 and α_2 may be used for these limit matrices and the correct relations may be found by their means among the quantities: trace, adjoint-trace, determinant and adjoint. It is merely necessary to regard α_1 and α_2 as not themselves in the domain of α, although commutative with α in multiplication and giving rise to the same characteristic functions. This is analogous to going from the Galois domains to non-Galois domains.

GENERALIZED CONJUGATE MATRICES.

By Philip Franklin.

The notion of conjugate matrices, which originated with O. Taber* and was applied in a recent paper by A. A. Bennett,† may be described in the following terms. If, corresponding to a given matrix M_1 of order n there exist $n - 1$ matrices of the same order satisfying the conditions:

1. They have the same characteristic equation as M_1;
2. They are commutative with respect to multiplication;
3. The symmetric functions ΣM_i, $\Sigma M_i M_j$, $\cdots$, $M_1 \cdot M_2 \cdot \cdots M_n$ formed from the matrix M_1 and these $n - 1$ matrices are scalars and equal to the corresponding functions of the n scalar roots of the characteristic equation of M_1;

these $n - 1$ matrices are called the $n - 1$ *conjugates* of M_1.

In the case where the roots of the characteristic equation of the given matrix are all distinct, the existence of the conjugate matrices is demonstrated‡ by noting that if r_1, r_2, $\cdots$ r_n are these n distinct roots, the matrix

$$(1) \qquad R_1 = \begin{vmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & r_n \end{vmatrix}$$

has the same elementary divisors as the given matrix M_1. Consequently§ a non-singular matrix P can be found such that:

$$(2) \qquad M_1 = PR_1P^{-1},$$

and the $n - 1$ matrices

$$(3) \qquad M_i = PR_iP^{-1},$$

where R_i is given by:

$$(4) \qquad R_i = \begin{vmatrix} r_i & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & r_{i+1} & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & r_n & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & r_1 & \cdots & 0 \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & 0 & 0 & \cdots & r_{i-1} \end{vmatrix},$$

evidently satisfy the three conditions stated above.

* Taber, O., On certain identities in the theory of matrices. Amer. Journ. Math., vol. 13 (1891), p. 159.

† Bennett, A. A., Some algebraic analogies in matric theory, these Annals, vol. 23, p. 91.

‡ Cf. Taber, l. c.

§ Böcher, M., Introduction to Higher Algebra, p. 283.

The same method applies to a matrix whose characteristic equation has equal roots, provided it has the same elementary divisors as a matrix of the form R_1, with some of the r's equal. In the case of matrices whose characteristic equations have equal roots, but whose elementary divisors are not of this type, the method fails. In fact in this case there may fail to exist a set of $n - 1$ matrices conjugate to the given matrix.*

The purpose of this note is to define a set of *generalized conjugate* matrices, which are subject to less stringent conditions than the ordinary conjugate matrices, but which have the advantage of existing in all cases. Furthermore, they are sufficient for many of the proofs given by Taber and Bennett which use ordinary conjugate matrices.

Our *generalized conjugate* matrices differ from the ordinary conjugate matrices in that they are required to satisfy conditions (2) and (3) only. The sacrifice of condition (1) is not as violent as it at first appears, since by the use of the remaining two conditions the characteristic equation:

$$M_1 - \lambda I = 0 \tag{5}$$

reduces to:

$$(M_1 - \lambda)(M_2 - \lambda) \cdots (M_n - \lambda) = 0 \tag{6}$$

and consequently the generalized conjugate matrices are *roots* of the characteristic equation.

To set up these matrices explicitly, we shall first notice† that any matrix is equivalent to a matrix of the form:

$$S_1 = \begin{vmatrix} S_1^1 & & & \\ & S_1^2 & & \\ & & \begin{matrix} \cdots \\ \cdots \\ \cdots \end{matrix} & \\ & & & S_1^p \end{vmatrix}, \tag{7}$$

where the missing elements are zeros and the S_1^i's are blocks of terms given by:

$$S_1^i = \begin{vmatrix} r_i & 1 & 0 & \cdots & 0 \\ 0 & r_i & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & r_i \end{vmatrix}. \tag{8}$$

Simple roots correspond to blocks of a single term. We therefore have, analogously to (2):

$$M_1 = PS_1P^{-1}. \tag{9}$$

* Bennett, l. c.

† Bôcher, l. c., p. 289.

In the case where S_1 consists of a single block,

(10) $$S_1 = \begin{Vmatrix} r_1 & 1 & 0 & \cdots & 0 \\ 0 & r_1 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & r_1 \end{Vmatrix},$$

the $n - 1$ matrices:

(11) $$M_i = PS_iP^{-1},$$

where S_i is given by

(12) $$S_i = \begin{Vmatrix} r_1 & \omega^{i-1} & 0 & \cdots & 0 \\ 0 & r_1 & \omega^{i-1} & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & r_1 \end{Vmatrix},$$

ω being a primitive nth root of unity, are evidently a set of $n - 1$ ordinary conjugates of M_1. In the general case, where M_1 satisfies the relations (9) and (7), we form the $n - 1$ matrices:

(13) $$S_i = \begin{Vmatrix} S_i^1 & & & \\ & S_i^2 & & \\ & & \begin{matrix} \cdots \\ \cdots \\ \cdots \end{matrix} & \\ & & & S_i^p \end{Vmatrix},$$

where for each block S_i^j of k rows, the $k - 1$ conjugates of S_1^j, the corresponding block of S_1, appear in $k - 1$ of the matrices, and in the remaining $n - k$ blocks its place is filled by a scalar block of the form:

(14) $$\begin{Vmatrix} r_i^{(j)} & 0 & \cdots & 0 \\ 0 & r_i^{(j)} & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & r_i^{(j)} \end{Vmatrix},$$

the r_i's being roots of the characteristic equation of M_1 and so selected that the $n - k$ values of r_i for the $n - k$ different values of i together with the k roots corresponding to S_1^j form the complete set of n roots of the characteristic equation of M_1. The generalized conjugate matrices of M_1 are then the $n - 1$ matrices obtained by combining (13) and (11). This is readily verified if we first notice that any rational integral function of the S_i's, $f(S_1, S_2, \cdots S_n)$, is given by:

$$(15)\quad f(S_1, S_2, \cdots S_n) =$$

$$\begin{Vmatrix} f(S_1^1, S_2^1, \cdots S_n^1) & & & \\ & f(S_1^2, S_2^2, \cdots S_n^2) & & \\ & & \begin{matrix} \cdots \\ \cdots \\ \cdots \end{matrix} & \\ & & & f(S_1^p, S_2^p, \cdots S_n^p) \end{Vmatrix}.$$

As an illustration, the matrix:

$$\begin{Vmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{Vmatrix}$$

has no matrices conjugate to it in the ordinary sense, but has as its generalized conjugates:

$$\begin{Vmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix} \quad \text{and} \quad \begin{Vmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{Vmatrix}.$$

It is to be noticed that the method of constructing the generalized conjugate matrices leads to a unique result only in exceptional cases, like that just given. For example, in addition to the ordinary conjugates, the matrix:

$$\begin{Vmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{Vmatrix}$$

has as generalized conjugates any of the pairs:

$$\begin{Vmatrix} b & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{Vmatrix}, \quad \begin{Vmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & a \end{Vmatrix};$$

$$\begin{Vmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & b \end{Vmatrix}, \quad \begin{Vmatrix} b & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{Vmatrix};$$

or:

$$\begin{Vmatrix} c & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{Vmatrix}, \quad \begin{Vmatrix} b & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & b \end{Vmatrix}.$$

PRINCETON UNIVERSITY.

TRANSFORMATIONS OF TRAJECTORIES ON A SURFACE.

By Joseph Lipka.

1. **Trajectories and their properties.** In a recent paper* the author proved five geometric properties which completely characterize the system of ∞^3 trajectories generated by the motion of a particle on any constraining surface under any positional field of force. The purpose of this paper is to study the point transformations on the surface which leave some or all of these properties invariant. Each of these properties together with those preceding it defines a type of systems of ∞^3 curves on the surface, and our problem is to find the nature of the transformations which convert any system of such a type into a system of the same type.† We shall here briefly state these properties, giving the differential equations of the systems of curves defined by them.

Let us consider the surface whose equations are

$$x = x(u, v), \qquad y = y(u, v), \qquad z = z(u, v),$$

referred to an orthogonal set of parameter curves,‡ so that the element of length has the form

$$ds^2 = Edu^2 + Gdv^2. \tag{1}$$

Property I. If the ∞^1 curves passing through a given point in a given direction have associated with them their orthogonal projections in the tangent plane to the surface at the given point, then the locus of the foci of the osculating parabolas of the associate system is a bicircular quartic with the given point as node and the given direction as tangent line; this tangent line is also one of the asymptotes to the hyperbola which is the inverse of the quartic with respect to the given point.

The most general system of ∞^3 curves on a surface possessing property I is defined by a differential equation of the form

* Motion on a surface for any positional field of force, Proc. Amer. Acad. Arts and Sci., vol. 56, no. 4, pp. 155–182. We shall hereafter refer to this paper by the title "Proceedings."

† For the corresponding problem in the plane, see E. Kasner, The trajectories of dynamics, Trans. Amer. Math. Soc., vol. 7, p. 418.

‡ In "Proceedings" we used an isothermal set of parameter curves, so that some of the equations had a simpler form. But for a discussion of point-transformations we find it necessary to use a more general parameter system.

$$(I) \qquad v''' = A + Bv'' + Cv''^2,$$

where A, B, C are arbitrary functions of u, v, v'.*

Property II. The two tangents at the node of the focal locus, associated with each element (u, v, v') by property I, are such that the one which has the direction of the given element bisects the angle between the other and a certain direction $\omega(u, v)$ through the given point (this direction is that of the force vector in the case of a trajectory system).

The most general system of ∞^3 curves on a surface possessing properties I and II is defined by a differential equation of the form

$$(II) \qquad v''' = A + Bv'' + \frac{3}{v' - \omega} v''^2,$$

where A and B are arbitrary functions of u, v, v', and ω is an arbitrary function of u, v.

Property III. Through every point and in every direction through that point there passes one curve of the system which hyperosculates its corresponding geodesic circle of curvature. The locus of the centers of geodesic curvature of the ∞^1 hyperosculating trajectories which pass through a point is a conic passing through the point in the direction $\omega(u, v)$ of property II.

The most general system of ∞^3 curves on a surface possessing properties I, II, III is defined by a differential equation of the form

$$(III) \qquad (\omega - v')H' = H(\gamma_0 + \gamma_1 v' + \gamma_2 v'^2 - 3v''),$$

involving four arbitrary functions, γ_0, γ_1, γ_2, ω of u, v, and where

$$(2) \quad H \equiv v'' - \frac{E_v}{2G} + \left(\frac{G_u}{G} - \frac{E_u}{2E}\right) v' + \left(\frac{G_v}{2G} - \frac{E_v}{E}\right) v'^2 + \frac{G_u}{2E} v'^3 = 0$$

is the differential equation of the geodesics on the surface, and

$$H' = dH/du.$$

Property IV. With each point O on the surface, property III associates a direction, viz., the tangent to the central locus or conic. The totality of all such directions on the surface defines a simple system of ∞^1 curves, which may be called the tangential lines (these are the lines of force in the trajectory system). The geodesic curvature of the tangential line through O is equal to 3 times the geodesic curvature of that hyperosculating curve which passes through O in the same direction.

The most general system of ∞^3 curves on a surface possessing properties I, II, III, IV is defined by a differential equation of the form (III) to-

* Throughout this paper, subscripts refer to partial derivatives with respect to the indicated variable and primes refer to total derivatives with respect to u.

gether with a condition on the functions γ_0, γ_1, γ_2, ω, i.e., by

$$(IV)\quad \begin{cases} (\omega - v')H' = H(\gamma_0 + \gamma_1 v' + \gamma_2 v'^2 - 3v'') \\ \gamma_0 + \gamma_1\omega + \gamma_2\omega^2 = (\omega_u + \omega\omega_v) \\ \quad - 2\left[-\dfrac{E_v}{2G} + \left(\dfrac{G_u}{G} - \dfrac{E_u}{2E}\right)\omega + \left(\dfrac{G_v}{2G} - \dfrac{E_v}{E}\right)\omega^2 + \dfrac{G_u}{2E}\omega^3 \right]. \end{cases}$$

Property V. Construct any isothermal net on the surface. At any point O this net determines two orthogonal directions in which there pass two isothermal curves of the net and two hyperosculating curves of property *III*. If ρ_1, ρ_2, R_1, R_2 are the radii of geodesic curvature of these four curves, s_1, s_2, the arc lengths along the isothermal curves, and ω, the tangent of the angle between the tangent line to the conic of property *III* and the isothermal curve with arc s_2, then as we move along the surface from O, these quantities vary so as to satisfy the relation

$$\frac{\partial}{\partial s_2}\left(\frac{1}{\kappa_1}\right) - \frac{\partial}{\partial s_1}\left(\frac{1}{\kappa_2}\right) - \frac{1}{\rho_1\kappa_1} + \frac{1}{\rho_2\kappa_2} - \frac{\partial^2}{\partial s_1 \partial s_2}(\log\omega) = 0,$$

where

$$\frac{1}{\kappa_1} = \omega\left(\frac{1}{\rho_1} - \frac{3}{R_1}\right); \qquad \frac{1}{\kappa_2} = \frac{1}{\omega}\left(\frac{1}{\rho_2} - \frac{3}{R_2}\right).$$

The most general system of ∞^3 curves on a surface possessing properties *I, II, III, IV, V* is defined by a differential equation of the form

$$(V)\qquad \left(\frac{\psi}{G} - \frac{\phi}{E}v'\right)H' = H\left(\delta_0 + \delta_1 v' + \delta_2 v'^2 - 3\frac{\phi}{E}v''\right),$$

where

$$(3)\quad \begin{cases} \delta_0 = \dfrac{\psi E_u + \phi E_v}{EG} + \dfrac{G\psi_u - \psi G_u}{G^2}, \quad \delta_2 = \dfrac{\phi E_v - E\phi_v}{E^2} - \dfrac{\psi G_u + \phi G_v}{EG}, \\ \delta_1 = \dfrac{2\psi E_v - 2\phi G_u}{EG} + \dfrac{G\psi_v - \psi G_v}{G^2} + \dfrac{\phi E_v - E\phi_u}{E^2}, \end{cases}$$

and where ϕ and ψ are arbitrary functions of u, v.

Equation (V) is the differential equation of the system of ∞^3 trajectories generated by a point moving on a surface under any positional field of force, the components of which along the tangent lines to the parameter curves are $\phi/\sqrt{E}$ and $\psi/\sqrt{G}$.* These determine a direction

* The trajectories on a surface are in general defined by the Lagrangian equations

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot u}\right) - \frac{\partial T}{\partial u} = \phi, \qquad \frac{d}{dt}\left(\frac{\partial T}{\partial \dot v}\right) - \frac{\partial T}{\partial v} = \psi,$$

where T is the kinetic energy, $\dot u = du/dt$, $\dot v = dv/dt$, $\phi = Xx_u + Yy_u + Zz_u$, $\psi = Xx_v + Yy_v + Zz_v$, X, Y, Z being the components of the force along the coördinate axes. See "Proceedings," § 2.

$\omega = dv/du = E\psi/G\phi$ along the surface, which we shall call the direction of the force vector.

2. Arbitrary point transformations. Consider any real representation of a surface S on a surface $\bar{S}$, whereby a one-to-one correspondence is established between the points of the two surfaces. In such an arbitrary point transformation there is at least one real orthogonal set of curves on S which corresponds to a real orthogonal set on $\bar{S}$.* If we choose the curves of these two orthogonal sets as parameter curves, and corresponding curves are assigned the same parameter value u or v, then corresponding points will have the same curvilinear coördinates u, v, and the elements of length are given by

$$(4) \qquad ds^2 = Edu^2 + Gdv^2, \qquad d\bar{s}^2 = \bar{E}du^2 + \bar{G}dv^2.$$

For the surface $\bar{S}$ we may set up the five types of differential equations $(\bar{I}), \cdots, (\bar{V})$, by merely writing $\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{\omega}$, $\bar{\gamma}_0$, $\bar{\gamma}_1$, $\bar{\gamma}_2$, $\bar{E}$, $\bar{G}$, $\bar{\phi}$, $\bar{\psi}$, for the corresponding letters in equations $(I), \cdots, (V)$. Now, if (I) is to be converted into $(\bar{I})$, it is necessary and sufficient that

$$(5) \qquad \bar{A} = A, \qquad \bar{B} = B, \qquad \bar{C} = C.$$

But since these coefficients are arbitrary functions of u, v, v', conditions (5) can always be satisfied. Similarly, if (II) is to be converted into $(\bar{II})$, it is necessary and sufficient that

$$(6) \qquad \bar{A} = A, \qquad \bar{B} = B, \qquad \bar{\omega} = \omega.$$

But since the coefficients are arbitrary functions, conditions (6) can always be satisfied. Hence we may state

THEOREM 1. *An arbitrary point transformation will convert any system of curves with property I or any system of curves with properties I and II on S into a like system on $\bar{S}$.*

3. Geodesic transformations. It is evident that an arbitrary point transformation will not convert (III) into $(\bar{III})$. To find the most general point transformation that will make this conversion, we note that such a transformation must convert the part common to all systems of type (III) into the part common to all systems of type $(\bar{III})$. It is evident that $H = 0$ and $\bar{H} = 0$ satisfy (III) and $(\bar{III})$ respectively, and that the curves defined by these equations are the only proper curves satisfying all equations of these types. But the curves $H = 0$ and $\bar{H} = 0$ are the geodesics on S and $\bar{S}$ respectively. Hence the desired transformation must convert the geodesics on S into the geodesics on $\bar{S}$. Such a

* See G. Scheffers, Anwendung der Differential- und Integral-Rechnung auf Geometrie, vol. 2, p. 96.

transformation is called a geodesic transformation. In order that the geodesics on S and $\bar{S}$ should correspond, it is necessary and sufficient that the differential equations $H = 0$ and $\bar{H} = 0$ be identical. We are thus led to the equations of condition

$$(7) \quad \begin{cases} -\dfrac{\bar{E}_v}{2\bar{G}} = -\dfrac{E_v}{2G}, & \dfrac{\bar{G}_u}{\bar{G}} - \dfrac{\bar{E}_u}{2\bar{E}} = \dfrac{G_u}{G} - \dfrac{E_u}{2E}, \\ \dfrac{\bar{G}_u}{2\bar{E}} = \dfrac{G_u}{2E}, & \dfrac{\bar{G}_v}{2\bar{G}} - \dfrac{\bar{E}_v}{\bar{E}} = \dfrac{G_v}{2G} - \dfrac{E_v}{E}, \end{cases}$$

which must hold identically. We note that this transformation will necessarily convert H into $\bar{H}$ and H' into $\bar{H}'$.

Now, if a geodesic transformation is to convert (III) into $(\overline{III})$, it is necessary and sufficient that

$$(8) \qquad \bar{\omega} = \omega, \qquad \bar{\gamma}_0 = \gamma_0, \qquad \bar{\gamma}_1 = \gamma_1, \qquad \bar{\gamma}_2 = \gamma_2.$$

Since these coefficients are arbitrary functions of u, v, conditions (8) can always be satisfied. Hence, we have

THEOREM 2. *The most general point transformation that converts any system of curves with properties I, II, III on S into a like system on $\bar{S}$, is the geodesic transformation.*

Now, equation (IV) involves the quantities E and G, which are not arbitrary. But it is evident that conditions (7) and (8) are both necessary and sufficient in order that (IV) be converted into $(\overline{IV})$. Hence, we have

THEOREM 3. *The most general point transformation that converts any system with properties I, II, III, IV on S into a like system on $\bar{S}$ is the geodesic transformation.*

Since type (V) is a special form of type (III), the most general point transformation that would convert any system of dynamical trajectories on S into a like system on $\bar{S}$ must be the geodesic transformation. We shall now examine whether every such transformation actually does convert a system of dynamical trajectories into a like system.

Dini first proved the theorem that if the real representation of a surface S on another $\bar{S}$ is geodesic, three cases are possible: (i) $\bar{S}$ may be obtained from S by a pure bending, i.e., $\bar{S}$ is applicable on S; (ii) $\bar{S}$ may be obtained from S by a similitude transformation with or without a bending; (iii) S and $\bar{S}$ are Liouville surfaces.* Let us apply these results to type (V).

(i) $\bar{S}$ is applicable on S; then

$$(9) \qquad \bar{E} = E, \qquad \bar{G} = G.$$

* See Scheffers, ibid., vol. 2, p. 420.

For (V) to be converted into $(\bar{V})$, corresponding coefficients must be equal. This leads to

$$\frac{\bar{\psi}}{\psi} = \frac{\bar{\phi}}{\phi} = \frac{\bar{\delta}_0}{\delta_0} = \frac{\bar{\delta}_1}{\delta_1} = \frac{\bar{\delta}_2}{\delta_2},$$

and these conditions reduce to

$$\frac{\bar{\psi}}{\psi} = \frac{\bar{\phi}}{\phi}, \qquad \frac{\bar{\psi}_u}{\bar{\psi}} = \frac{\psi_u}{\psi}, \qquad \frac{\bar{\phi}_v}{\bar{\phi}} = \frac{\phi_v}{\phi}, \qquad \frac{\bar{\psi}_v}{\bar{\psi}} - \frac{\bar{\phi}_u}{\bar{\psi}} = \frac{\psi_v}{\psi} - \frac{\phi_u}{\psi},$$

and hence,

$$(10) \qquad \bar{\psi} = c\psi, \qquad \bar{\phi} = c\phi,$$

where c is an arbitrary constant. We may take $c = 1$, since ϕ, ψ and $c\phi$, $c\psi$ define the same field of force. Conditions (10) can always be satisfied since $\bar{\phi}$, $\bar{\psi}$ are arbitrary functions of u, v.

(ii) $\bar{S}$ is obtained from S by a similitude transformation, i.e., the rectangular coördinates of corresponding points on the two surfaces are connected by the relations

$$\bar{x} = kx, \qquad \bar{y} = ky, \qquad \bar{z} = kz, \qquad (k = \text{constant});$$

then

$$(11) \qquad \bar{E} = k^2E, \qquad \bar{G} = k^2G.$$

Again, if (V) is to be converted into $(\bar{V})$, then, as in case *(i)*, $\bar{\psi} = \psi$, $\bar{\phi} = \phi$. Hence, we have

THEOREM 4. *Any system of dynamical trajectories on S is converted by a geodesic transformation into a system of dynamical trajectories on $\bar{S}$, provided $\bar{S}$ is applicable on S or $\bar{S}$ is applicable on a surface which can be obtained from S by a similitude transformation. The components of force on S and $\bar{S}$ are the same functions of the coördinates u, v.*

(iii) S and $\bar{S}$ are Liouville surfaces. A Liouville surface is characterized by the fact that the element of length may be reduced to the form

$$(12) \qquad ds^2 = (U + V)(du^2 + dv^2),$$

where U is a function of u only and V a function of v only.* The corresponding surface $\bar{S}$ has for element of length

$$(13) \qquad d\bar{s}^2 = -\left(\frac{1}{U} + \frac{1}{V}\right)\left(\frac{du^2}{U} - \frac{dv^2}{V}\right).$$

* To the Liouville surfaces belong, among others, the surfaces of constant curvature, the surfaces of revolution, and the quadrics. When the element of length can be written in the form (12), the parameter curves are geodesic ellipses and hyperbolas. Cf. L. P. Eisenhart, Differential Geometry, p. 215. For Liouville surfaces the finite equation of the geodesics may be found by simple quadratures. For a discussion of surfaces of this type, see Darboux, Leçons sur la Théorie Générale des Surfaces, vol. II, Chap. IX.

It is easily seen that $d\bar{s}^2$ has the Liouville form, for by a change of parameters

$$\bar{u} = \int \frac{du}{\sqrt{U}}, \qquad \bar{v} = \int \frac{dv}{\sqrt{-V}},$$

(13) takes the form

$$d\bar{s}^2 = -\left(\frac{1}{U} + \frac{1}{V}\right)(d\bar{u}^2 + d\bar{v}^2) = (\bar{U} + \bar{V})(d\bar{u}^2 + d\bar{v}^2).$$

Now, on the surface S defined by (12), type (V) takes the form

(14) $$(\psi - \phi v')H' = H(a_1 + a_2 v' + a_3 v'^2 - 3\phi v''),$$

where

$$a_1 = \psi_u + \frac{\phi V_v}{U + V}, \qquad a_2 = \psi_v - \phi_u + \frac{\psi V_v - \phi U_u}{U + V},$$

$$a_3 = -\phi_v - \frac{\psi U_u}{U + V}.$$

On the surface $\bar{S}$ defined by (13), type $(\bar{V})$ takes the form

(15) $$\left(-\frac{V\bar{\psi}}{U} - \bar{\phi}v'\right)H' = H(b_1 + b_2 v' + b_3 v'^2 - 3\bar{\phi}v''),$$

where

$$b_1 = \frac{V(\bar{\psi}U_u - U\bar{\psi}_u)}{U^2} + \frac{\bar{\phi}V_v}{U + V}, \quad b_3 = \frac{\bar{\phi}V_v - V\bar{\phi}_v}{V} + \frac{\bar{\psi}VU_u}{U(U + V)},$$

$$b_2 = -\frac{V\bar{\psi}_v}{U} - \bar{\phi}_u - \frac{\bar{\psi}VV_v + \bar{\phi}UU_u}{U(U + V)}.$$

Here ϕ, ψ and $\bar{\phi}$, $\bar{\psi}$ are arbitrary functions of u, v determining the force vectors on S and $\bar{S}$ respectively. Let us now see whether by a proper choice of $\bar{\phi}$ and $\bar{\psi}$, equation (14) may be converted into equation (15). For such conversion, corresponding coefficients must be equal, i.e.,

(16) $$\frac{\psi}{\phi} = -\frac{V\bar{\psi}}{U\bar{\phi}}, \qquad \frac{a_1}{\phi} = \frac{b_1}{\bar{\phi}}, \qquad \frac{a_2}{\phi} = \frac{b_2}{\bar{\phi}}, \qquad \frac{a_3}{\phi} = \frac{b_3}{\bar{\phi}}.$$

Combining the first two conditions, we find

$$\frac{\bar{\psi}_u}{\bar{\psi}} = \frac{U_u}{U} + \frac{\psi_u}{\psi}.$$

Hence

(17) $$\bar{\psi} = V_1 U\psi, \qquad \bar{\phi} = -V_1 V\phi,$$

where V_1 is an arbitrary function of v only. Combining the first and

third condition, we find

$$\frac{\bar{\phi}_v}{\bar{\phi}} = \frac{V_v}{V} + \frac{\phi_v}{\phi},$$

and substituting the value of $\bar{\phi}$ found in (17), this becomes

$$V_1 = \text{constant} = c.$$

Hence,

$$\bar{\psi} = cU\psi, \qquad \bar{\phi} = -cV\phi. \tag{18}$$

By substitution we find that these relations satisfy the fourth condition. As in case (*i*) we may take $c = 1$, so that

$$\bar{\psi} = U\psi, \qquad \bar{\phi} = -V\phi. \tag{19}$$

These conditions can always be satisfied, since $\bar{\phi}$, $\bar{\psi}$ are arbitrary functions of u, v. Hence, we have

THEOREM 5. *A system of dynamical trajectories on a Liouville surface S is converted by a geodesic transformation into a like system on the corresponding Liouville surface $\bar{S}$. The components of force ϕ, ψ on S and the components $\bar{\phi}$, $\bar{\psi}$ on $\bar{S}$ are related by $\bar{\psi} = U\psi$, $\bar{\phi} = -V\phi$.*

Finally, combining Theorems 4 and 5, we may state

THEOREM 6. *Any geodesic transformation of a surface S into a surface $\bar{S}$ will convert any system of dynamical trajectories on S into a like system on $\bar{S}$; a geodesic transformation is the most general point transformation that makes this conversion possible.*

4. Conservative forces. If the field of force is conservative, then $\psi_u = \phi_v$. This condition is characterized geometrically by the fact that the conic of property *III* becomes a rectangular hyperbola.* The question arises: is a system of dynamical trajectories in a conservative field of force converted by a geodesic transformation into a like system? This may be answered in the affirmative for cases (*i*) and (*ii*) in our discussion of geodesic transformations, for, by Theorem 4, $\bar{\psi} = \psi$, $\bar{\phi} = \phi$; hence, if $\psi_u = \phi_v$, then $\bar{\psi}_u = \bar{\phi}_v$. But for case (*iii*), where S and $\bar{S}$ are Liouville surfaces, we have, by Theorem 5, $\bar{\psi} = U\psi$, $\bar{\phi} = -V\phi$. Now if $\psi_u = \phi_v$ and $\bar{\psi}_u = \bar{\phi}_v$, we must have

$$(U + V)\phi_v + U_u\psi + V_v\phi = 0. \tag{20}$$

Now, if W is the work function (negative potential), then

$$\phi = W_u, \qquad \psi = W_v$$

and (20) becomes the Laplacian equation

* "Proceedings," § 9.

$$(U + V)W_{uv} + V_vW_u + U_uW_v = 0, \tag{21}$$

or

$$[(U + V)W]_{uv} = 0, \tag{22}$$

the solution of which is

$$(U + V)W = U_1 + V_1, \tag{23}$$

where U_1 is an arbitrary function of u alone, and V_1 is an arbitrary function of v alone. Hence, we may state

THEOREM 7. *A geodesic transformation will convert a system of dynamical trajectories in a conservative field of force on any surface S into a like system on $\bar{S}$, provided $\bar{S}$ is applicable on S, or $\bar{S}$ is applicable on a surface which may be obtained from S by a similitude transformation. A geodesic transformation will convert a system of dynamical trajectories in a conservative field of force on a Liouville surface S into a like system on the corresponding Liouville surface $\bar{S}$ only if W, the work function, has the form*

$$W = (U_1 + V_1)/(U + V),$$

where U_1 and V_1 are arbitrary functions of u and v respectively.

From equation (23) it is evident that there are no Liouville surfaces for which the transformation is possible if W is to be an arbitrary point function.

5. **"N" systems.** If the field of force is conservative, we may study the transformations of certain types of ∞^3 curves on a surface other than dynamical trajectories. These systems, termed "n" systems,* are characterized by the differential equation

$$(V_n) \quad \begin{cases} \left(\dfrac{\psi}{G} - \dfrac{\phi}{E}v'\right)(E + Gv'^2)H' \\ \qquad = H\Big\{(E + Gv'^2)\left(\epsilon_0 + \epsilon_1 v' + \epsilon_2 v'^2 - \dfrac{3\phi}{E}v''\right) \\ \qquad + \dfrac{(n-2)v'}{2}(\zeta_0 + \zeta_1 v' + \zeta_2 v'^2) - (n-2)(\phi + \psi v')v''\Big\} \end{cases}$$

where

$$\epsilon_0 = \frac{2\psi E_u + n\phi E_v}{2EG} + \frac{G\psi_u - \psi G_u}{G^2}, \qquad \epsilon_2 = -\frac{n\psi G_u + 2\phi G_v}{2EG} + \frac{\phi E_v - E\phi_v}{E^2},$$

$$\epsilon_1 = \frac{(2+n)(\psi E_v - \phi G_u)}{2EG} + \frac{G\psi_v - \psi G_v}{G^2} + \frac{\phi E_u - E\phi_u}{E^2}$$

and

$$\zeta_0 = \frac{\phi E_u}{E} - \frac{\phi G_u}{G}, \quad \zeta_2 = \frac{\psi E_v}{E} - \frac{\psi G_v}{G}, \quad \zeta_1 = \frac{\psi E_u}{E} - \frac{\psi G_u}{G} + \frac{\phi E_v}{E} - \frac{\phi G_v}{G}$$

and where $\phi_v = \psi_u$.

* "Proceedings," § 10.

An "n" system is a system of dynamical trajectories when $n = 2$, velocity curves when $n = 0$, brachistochrones when $n = -2$, catenaries when $n = 1$. Even if the field is not conservative, equation (V_n) may still be said to define "n" systems—dynamical trajectories, velocity curves, pseudo-brachistochrones, pseudo-catenaries.

In "Proceedings" we have given five geometric properties—I_n, II_n, III_n, IV_n, V_n—which completely characterize such "n" systems. These properties are analogous to properties $I, \cdots, V$ which characterize the trajectories under any positional field of force. In fact, I and I_n are the same; to get II_n we replace the equal angles in II by two angles whose tangents are in the ratio $n + 1 : 3$; III and III_n are the same—for the conservative case the conic must be a rectangular hyperbola; to get IV_n and V_n, we replace the multiple 3 in IV and V by $n + 1$. Each of these properties together with the preceding may be characterized by differential equations $(I_n), \cdots, (V_n)$ similar to equations $(I), \cdots, (V)$.

Equation (V_n) is given above. The others are

$$(I_n) \qquad v''' = A + Bv'' + Cv''^2,$$

$$(II_n) \qquad v''' = A + Bv'' + \frac{1}{\omega - v'}\left[\frac{(2-n)(E + G\omega v')}{E + Gv'^2} - 3\right]v''^2,$$

$$(III_n) \qquad (\omega - v')H' = H\left\{\gamma_0 + \gamma_1 v' + \gamma_2 v'^2 + \left[\frac{(2-n)(E + G\omega v')}{E + Gv'^2} - 3\right]v''\right\},$$

$$(IV_n) \qquad \begin{cases} (\omega - v')H' = H\left\{\gamma_0 + \gamma_1 v' + \gamma_2 v'^2 + \left[\dfrac{(2-n)(E + G\omega v')}{E + Gv'^2} - 3\right]v''\right\}, \\ \gamma_0 + \gamma_1\omega + \gamma_2\omega^2 = (\omega_u + \omega\omega_v) - n\left\{-\dfrac{E_v}{2G} + \left(\dfrac{G_u}{G} - \dfrac{E_u}{2E}\right)\omega + \left(\dfrac{G_v}{2G} - \dfrac{E_v}{E}\right)\omega^2 + \dfrac{G_u}{2E}\omega^3\right\}. \end{cases}$$

The quantities $A, B, C, \omega, \gamma_0, \gamma_1, \gamma_2, \psi, \phi, H, H'$ have the same significance as in the previous discussion.

A study of equations $(I_n), \cdots, (V_n)$ similar to the study made of equations $(I), \cdots, (V)$ leads to the following results:

1. Type (I_n) alone is conserved under an *arbitrary* point transformation.

2. In order that (II_n) be converted into $(\overline{II}_n)$, it is necessary and sufficient that $\overline{E}/E = \overline{G}/G$, i.e., the transformation must be conformal.

3. In order that (III_n) be converted into $(\overline{III}_n)$, the transformation must be *geodesic*, but not every such transformation makes the desired conversion. Under case (i), $\overline{E} = E, \overline{G} = G$; and under case (ii), $\overline{E} = k^2E$, $\overline{G} = k^2G$. These types of geodesic transformation evidently convert (III_n) into $(\overline{III}_n)$. But under case (iii), i.e., on Liouville surfaces, where $E = G = U + V$, and

$$\overline{E} = -\left(\frac{1}{U}+\frac{1}{V}\right)\frac{1}{U}, \qquad \overline{G} = \left(\frac{1}{U}+\frac{1}{V}\right)\frac{1}{V},$$

(III_n) is converted into $(\overline{III}_n)$ if, and only if, $U + V = 0$, i.e., the element of length ds vanishes over the entire surface. Hence, a geodesic transformation on a Liouville surface will not make the desired conversion.

4. A geodesic transformation of types (i) and (ii) will convert (IV_n) into $(\overline{IV}_n)$ and (V_n) into $(\overline{V}_n)$ with $\overline{\psi} = \psi$, $\overline{\phi} = \phi$.

Of course, the more general results of the earlier discussion hold for the case $n = 2$, i.e., dynamical trajectories.

We may now state

THEOREM 8. *An arbitrary point transformation will convert any system of curves with property I_n on S into a like system on $\overline{S}$. The most general point transformation that will convert any system of curves with properties I_n, II_n on S into a like system on $\overline{S}$ is the conformal transformation. The most general point transformation that will convert systems of curves with properties I_n, II_n, III_n, or I_n, II_n, III_n, IV_n, or I_n, II_n, III_n, IV_n, V_n (i.e., an " n " system) on S into a like system on $\overline{S}$ is the geodesic transformation under which $\overline{S}$ is applicable on S or $\overline{S}$ is applicable on a surface which can be obtained from S by a similitude transformation. Conservative fields are converted into conservative fields.*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
CAMBRIDGE, MASS.,
November, 1920.

ON THE STRUCTURE OF FINITE CONTINUOUS GROUPS WITH ONE TWO-PARAMETER INVARIANT SUBGROUP.

BY S. D. ZELDIN.

In a paper published in these Annals* I have considered groups having exceptional transformations and have shown how their structure can be simplified by imposing certain conditions on groups isomorphic with the given ones. In the present paper I shall show how the structure of groups with *one two-parameter invariant subgroup* can be simplified by imposing a few conditions on the groups *meroëdrically isomorphic* with them.

1. **Introductory remarks and assumptions.** Let G_{r+2} be a finite continuous group of order $r+2$ generated by the infinitesimal transformations whose differential operators are $X_1, \cdots, X_r, X_{r+1}, X_{r+2}$, where

$$X_i = \sum_{k=1}^{r+2} \xi(x_1, \cdots, x_{r+2}) \frac{\partial}{\partial x_k} \qquad (i = 1, \cdots, r+2),$$

and let G_{r+2} have an invariant two-parameter subgroup, which for simplicity may be taken to be generated by the operators X_{r+1} and X_{r+2}. Denoting the operators of the adjoint of G_{r+2} by $E_1, \cdots, E_{r+2}$, where

$$E_i = \sum_{j=1}^{r+2} \sum_{k=1}^{r+2} \alpha_j c_{jik} \frac{\partial}{\partial \alpha_k} \qquad (i = 1, \cdots, r+2)$$

(the α's are the parameters of G_{r+2} and the c_{jik}'s are the structural constants), we can write down the following known equalities:

$$(1) \qquad (X_i, X_j) = \sum_{k=1}^{r} c_{ijk} X_k + \sum_{k=r+1}^{r} c_{ijk} X_k \qquad (i, j = 1, \cdots, r),$$

$$(2) \qquad (X_i, X_j) = \sum_{k=r+1}^{r+2} c_{ijk} X_k \qquad \begin{pmatrix} i = 1, \cdots, r+1, r+2 \\ j = r+1, r+2 \end{pmatrix},$$

$$(3) \qquad (E_i, E_j) = \sum_{k=1}^{r+2} c_{ijk} E_k \qquad (i, j = 1, \cdots, r).$$

Since the group G_{r+2} is assumed to have an invariant subgroup of order 2, there exists a *simple* group of order r, say G_r, which is meroëdrically isomorphic with G_{r+2}.† If we denote the operators of G_r by $Y_1, \cdots, Y_r$

* Vol. 22, p. 95.

† Lie-Engel, vol. 3, p. 703.

where

$$Y_i = \sum_{k=1}^{r} \theta_{ki}(y_1, \cdots, y_r) \frac{\partial}{\partial y_k},$$

and the operators of the adjoint of G_r by $A_1, \cdots, A_r$, where

$$A_i = \sum_{k=1}^{r} \sum_{j=1}^{r} \alpha_j c_{jik} \frac{\partial}{\partial \alpha_k},$$

we have

$$(Y_i, Y_j) = \sum_{k=1}^{r} c_{ijk} Y_k \qquad (i, j = 1, \cdots, r)$$

and

$$(A_i, A_j) = \sum_{k=1}^{r} c_{ijk} A_k \qquad (i, j = 1, \cdots, r).$$

The condition imposed on the adjoint of G_r is that it shall have *one invariant spread.* It is to be observed that this spread is not a flat, for if it were, the group G_r would not be simple. If the invariant spread is given by the equation

$$F(\alpha_1, \cdots, \alpha_r) = 0,$$

then the function $F(\alpha_1, \cdots, \alpha_r)$ will satisfy the system of partial differential equations

$$A_i f(\alpha_1, \cdots, \alpha_r) \equiv \Sigma \alpha_j c_{ji1} \frac{\partial f}{\partial \alpha_1} + \cdots + \Sigma \alpha_j c_{jir} \frac{\partial f}{\partial \alpha_r} = 0 \quad (i = 1, \cdots, r).$$

Forming the matrix of the coefficients of those differential equations

$$\sum_{j=1}^{r} \alpha_j A_j \equiv \begin{pmatrix} \Sigma \alpha_j c_{j11}, & \Sigma \alpha_j c_{j12}, & \cdots & \Sigma \alpha_j c_{j1r} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \Sigma \alpha_j c_{jr1}, & \Sigma \alpha_j c_{jr2}, & \cdots & \Sigma \alpha_j c_{jrr} \end{pmatrix}$$

we must have, since that system of equations has only one solution, the nullity of this matrix to be equal to *one*, i.e., at least one minor of order $r-2$ of the determinant $|\Sigma \alpha_j A_j|$ does not vanish. But each minor of $|\Sigma \alpha_j A_j|$ is also a minor of $|\Sigma \alpha_j E_j|$, where

$$\sum_{j=1}^{r+2} \alpha_j E_j \equiv \begin{pmatrix} \Sigma \alpha_j c_{j,1,1}, & \cdots & \Sigma \alpha_j c_{j,r,1}, & \Sigma \alpha_j c_{j,r+1,1}, & \Sigma \alpha_j c_{j,r+2,1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \Sigma \alpha_j c_{j,1,r}, & \cdots & \Sigma \alpha_j c_{j,r,r}, & \Sigma \alpha_j c_{j,r+1,r}, & \Sigma \alpha_j c_{j,r+2,r} \\ \Sigma \alpha_j c_{j,1,r+1}, & \cdots & \Sigma \alpha_j c_{j,r,r+1}, & \Sigma \alpha_j c_{j,r+1,r+1}, & \Sigma \alpha_j c_{j,r+2,r+1} \\ \Sigma \alpha_j c_{j,1,r+2}, & \cdots & \Sigma \alpha_j c_{j,r,r+2}, & \Sigma \alpha_j c_{j,r+1,r+2}, & \Sigma \alpha_j c_{j,r+2,r+2} \end{pmatrix}.$$

Therefore at least one minor of order $r - 3$ of the determinant $|\Sigma\alpha_j E_j|$ does not vanish, and thus the nullity of the matrix $\Sigma\alpha_j E_j$ can not exceed 3 for an arbitrary system of values of the α's. Further, for $\alpha_1, \cdots \alpha_{r+2}$ assigned, the symbolic equation

$$(A) \qquad \left(\sum_{i=1}^{r+2}\alpha_i X_i, \sum_{j=1}^{r+2}\beta_j X_j\right) = \rho\Sigma\Sigma\alpha_i\beta_j \sum_{k=r+1}^{r+2} c_{ijk}X_k \qquad (\rho \neq 0)$$

is satisfied by the following three independent solutions:

$$\begin{array}{llllllr} \beta_1 = \alpha_1, & \beta_2 = \alpha_2, & \cdots, & \beta_r = \alpha_r, & \beta_{r+1} = 0, & \beta_{r+2} = 0, & (1) \\ \beta_1 = 0, & \beta_2 = 0, & \cdots, & \beta_r = 0, & \beta_{r+1} = 1, & \beta_{r+2} = 0, & (2) \\ \beta_1 = 0, & \beta_2 = 0, & \cdots, & \beta_r = 0, & \beta_{r+1} = 0, & \beta_{r+2} = 1. & (3) \end{array}$$

The first set of β's satisfies, because it makes both sides of equation (A) equal to zero; the sets (2) and (3) satisfy because X_{r+1} and X_{r+2} form, by our assumption, an invariant subgroup of G_{r+2}.*

But from equation (A) follows the system of equations

$$\beta_1\sum_{i=1}^{r+2}\alpha_i c_{i11} + \beta_2\sum_{i=1}^{r+2}\alpha_i c_{i21} + \cdots + \beta_{r+1}\sum_{i=1}^{r+2}\alpha_i c_{i,\,r+1,\,1} + \beta_{r+2}\sum_{i=1}^{r+2}\alpha_i c_{i,\,r+2,\,1} = 0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\beta_1\sum_{i=1}^{r+2}\alpha_i c_{i1r} + \beta_2\sum_{i=1}^{r+2}\alpha_i c_{i2r} + \cdots + \beta_{r+1}\sum_{i=1}^{r+2}\alpha_i c_{i,\,r+1,\,r} + \beta_{r+2}\sum_{i=1}^{r+2}\alpha_i c_{i,\,r+2,\,r} = 0,$$

$$(1-\rho)\Big[\beta_1\sum_{i=1}^{r+2}\alpha_i c_{i,\,1,\,r+1} + \beta_2\sum_{i=1}^{r+2}\alpha_i c_{i,\,2,\,r+1} + \cdots$$
$$+ \beta_{r+1}\sum_{i=1}^{r+2}\alpha_i c_{i,\,r+1,\,r+1} + \beta_{r+2}\sum_{i=1}^{r+2}\alpha_i c_{i,\,r+2,\,r+2}\Big] = 0,$$

$$(1-\rho)\Big[\beta_1\sum_{i=1}^{r+2}\alpha_i c_{i,\,1,\,r+2} + \beta_2\sum_{i=1}^{r+2}\alpha_i c_{i,\,2,\,r+2} + \cdots$$
$$+ \beta_{r+1}\sum\alpha_i c_{i,\,r+1,\,r+2} + \beta_{r+2}\sum_{i=1}^{r+2}\alpha_i c_{i,\,r+2,\,r+2}\Big] = 0.$$

The determinant of the coefficients of the β's in those equations must have at least one non-vanishing minor of order $r - 1$, and therefore the nullity of the matrix $\sum_{j=1}^{r+2}\alpha_j E_j$ can not be less than 3. We may now say that the nullity of $\sum_{j=1}^{r+2}\alpha_j E_j$ is *equal to 3*. Now, since the nullity of the matrix $\Sigma\alpha_j E_j$ is equal to the number of independent invariants of the adjoint of G_{r+2}, the system of partial differential equations

$$E_1 f(\alpha) = 0, \qquad \cdots, \qquad E_{r+2} f(\alpha) = 0$$

* Since all the operators of the group G_{r+2} are given, ρ and c_{ijk} $(k = r+1, r+2)$ can easily be found.

has three independent functions in $\alpha_1, \cdots, \alpha_{r+2}$ for solutions. It is also evident that $F(\alpha_1, \cdots, \alpha_r)$, which is a solution of the equations

$$A_1f(\alpha) = 0, \qquad \cdots, \qquad A_rf(\alpha) = 0,$$

is also a solution of

$$E_1f(\alpha) = 0, \qquad \cdots, \qquad E_{r+2}f(\alpha) = 0,$$

for $F(\alpha)$ does not depend on $\alpha_{r+1}, \alpha_{r+2}$. Denoting the invariants of the adjoint of G_{r+2} by $F(\alpha_1, \cdots, \alpha_r)$, $V(\alpha_1, \cdots, \alpha_{r+2})$, $W(\alpha_1, \cdots, \alpha_{r+2})$, we may state the following

THEOREM. *If the adjoint of G_r, which is meroëdrically isomorphic with G_{r+2}, has one invariant, the adjoint of G_{r+2} has three invariants, one of which is also invariant to the adjoint of G_r.*

2. **The invariant spreads of the adjoint of G_{r+2} and their properties.** Consider the invariant spread $V(\alpha_1, \cdots, \alpha_{r+2}) = 0$, supposing that it is the only $(r + 1)$-flat invariant to the adjoint of G_{r+2}. It will then represent an invariant subgroup of order $r + 1$ of G_{r+2}.* It is to be observed that the two-parameter subgroup X_{r+1}, X_{r+2} which was assumed to be invariant in G_{r+2} represents geometrically a straight-line invariant in the space of the adjoint of G_{r+2}. We shall denote, in what follows, that line by the symbol $X_{r+1} \leftarrow\rightarrow X_{r+2}$. Now, if the invariant flat $V(\alpha_1, \cdots, \alpha_{r+2}) = 0$ does not pass through the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$, there will then be in G_{r+2} an invariant subgroup of order $r + 1$ in addition to the given two-parameter invariant subgroup. In other words, we can find a new set of operators $\overline{X}_1, \cdots, \overline{X}_{r+1}, \overline{X}_{r+2}$, such linear functions of the old X's, that

$$(\overline{X}_i, \overline{X}_j) = \sum_{k=1}^{r+1} c_{ijk}X_k \quad \begin{pmatrix} i = 1, \cdots, r+1, r+2 \\ j = 1, \cdots, r+1 \end{pmatrix}.$$

If however the flat $V(\alpha_1, \cdots, \alpha_{r+2}) = 0$ does pass through the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$, then the point of intersection would have to be invariant to adjoint of G_{r+2} and there would be in G_{r+2} an invariant subgroup of order one. This case brings us to exceptional transformations which I have already discussed in my last paper.

Suppose now that $V(\alpha_1, \cdots, \alpha_{r+2}) = 0$ is an equation of degree two, reducible to two linear equations, say $V_1(\alpha_1, \cdots, \alpha_{r+2}) = 0$ and $V_2(\alpha_1, \cdots, \alpha_{r+2}) = 0$, then the intersection of these two $(r + 1)$-flats will give an invariant r-flat in the space of the adjoint of G_{r+2}. If this r-flat does not pass through the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$, we will be able to find r independent operators forming an invariant subgroup of order r of G_{r+2}.

* Lie-Scheffers, p. 479.

An interesting case will arise when $V(\alpha_1, \cdots, \alpha_{r+2}) = 0$ is an irreducible algebraic spread of degree $m \geqq 2$. Let us consider an arbitrary point P on the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$. Its polar $(r+1)$-flat, with respect to the spread $V(\alpha_1, \cdots, \alpha_{r+2}) = 0$, will in general pass through some other point, say Q, of the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$, and the polar $(r+1)$-flat of Q, with respect to the same spread, will then pass through P. The intersection of those two $(r+1)$-flats will give an r-flat which may be regarded as a polar r-flat, with respect to $V(\alpha_1, \cdots, \alpha_{r+2}) = 0$, of the line PQ (or $X_{r+1} \leftarrow\rightarrow X_{r+2}$). That flat may be looked upon as the locus of the poles of all $(r+1)$-flats passing through the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$ taken with respect to the spread $V(\alpha_1, \cdots, \alpha_{r+2}) = 0$.*

Now, since the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$ and the spread $V(\alpha) = 0$ are invariant to the adjoint of G_{r+2}, the aggregate of flats passing through the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$ and therefore the locus of their poles quâ $V(\alpha) = 0$ will be invariant. Since that locus of poles is an r-flat, it follows that the group G_{r+2} has *an invariant subgroup of order* r, i.e., by properly choosing the operators $\overline{X}_1, \cdots, \overline{X}_{r+2}$ we shall have

$$(\overline{X}_i, \overline{X}_j) = \sum_{k=1}^{r} \bar{c}_{ijk} X_k \quad \begin{pmatrix} i = 1, \cdots, r, r+1, r+2 \\ j = 1, \cdots, r \end{pmatrix}.$$

Suppose, however, that $V(\alpha) = 0$ is of degree m, but is reducible to m $(r+1)$-flats. Then their common intersection (if there is any) will form an invariant $(r-m+2)$-flat. If that flat does not pass through the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$, then there will be in G_{r+2} an *invariant subgroup of order* $r - m + 2$ *in addition to the given two-parameter subgroup*, i.e., the operators

$$\overline{X}_1, \cdots, \overline{X}_{r-m+2}, \overline{X}_{r-m+1}, \cdots, \overline{X}_r, \overline{X}_{r+1}, \overline{X}_{r+2}$$

can be so chosen that

$$(\overline{X}_i, \overline{X}_j) = \sum_{k=1}^{r-m+2} \bar{c}_{ijk} X_k \qquad \begin{pmatrix} i = 1, \cdots, r-m+2, \cdots, r \\ j = 1, \cdots, r-m+2 \end{pmatrix}$$

$$(X_i, X_{r+1}) = (X_i, X_{r+2}) = 0 \qquad (i = 1, \cdots, r-m+2)$$

$$(X_i, X_j) = \sum_{r+1}^{r+2} \bar{c}_{ijk} X_k \qquad \begin{pmatrix} i = r-m+2, \cdots, r+1, r+2 \\ j = \qquad\qquad r+1, r+2 \end{pmatrix}.$$

If the $(r-m+2)$-flat does pass through the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$, we again get a single invariant point in the space of the adjoint of G_{r+2}, whose meaning I discussed before.

It may happen that $V(\alpha) = 0$ breaks up into spreads each of degree greater than one. Then their common intersection and its polar flat,

* Compare with Salmon's discussion of polar lines, G. Salmon, Analytic Geometry of Three Dimensions, p. 49.

taken with respect to the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$, can be considered exactly in the same way as when $V(\alpha) = 0$ was irreducible.

So far I have only considered the invariant $V(\alpha)$ independently of the third invariant $W(\alpha)$ of the adjoint of G. We could of course obtain the same results for $W(\alpha)$, as we did for $V(\alpha)$, by considering it alone. Suppose, however, that the invariant spreads

$$V(\alpha) = 0 \qquad \text{and} \qquad W(\alpha) = 0$$

are taken together, and assume first that both are $(r + 1)$-flats. If their intersection, which is an r-flat, does not pass through the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$, then there is an invariant subgroup of order r of G_{r+2} in addition to the given invariant two-parameter subgroup. If however the $(r + 1)$-flats do not intersect at all, then each one separately will represent an invariant subgroup of order $r + 1$ of G_{r+2}.

If finally $V(\alpha) = 0$ and $W(\alpha) = 0$ are spreads of degrees m and n respectively, then, by considering the polar flat of the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$, taken with respect to the intersection of $V(\alpha) = 0$ and $W(\alpha) = 0$, we shall get an invariant subgroup of order r of G_{r+2}.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
November 8, 1920.

ON THE SIMPLIFICATION OF THE STRUCTURE OF FINITE CONTINUOUS GROUPS WITH MORE THAN ONE TWO-PARAMETER INVARIANT SUBGROUP.

By S. D. Zeldin.

1. It is the purpose of this paper to extend the results obtained for the structure of groups with one two-parameter invariant subgroup* to groups having any number of two-parameter invariant subgroups.

Let $X_1, \cdots, X_r, X_{r+1}, \cdots, X_{r+2k}$ be the operators of the group G_{r+2k} whose order is $r + 2k$. Let this group have k invariant two-parameter subgroups which for simplicity will be taken to be represented by the operators

$$\begin{array}{lll} X_{r+1}, & X_{r+2} & (1), \\ X_{r+3}, & X_{r+4} & (2), \\ X_{r+5}, & X_{r+6} & (3), \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot \\ X_{r+2k-1}, & X_{r+2k} & (k). \end{array}$$

We then have

$$(X_i, X_j) = \sum_{s=1}^{r+2k} c_{ijs} X_s \qquad (i, j = 1, 2, \cdots, r),$$

$$(X_i, X_j) = \rho_{ij} \sum_{s=r+1}^{r+2} c_{ijs} X_s \quad (i = 1, 2, \cdots, r;\ j = r+1, r+2;\ \rho_{ij} \neq 0),$$

$$(X_i, X_j) = \rho_{ij} \sum_{s=r+3}^{r+4} c_{ijs} X_s \quad (i = 1, 2, \cdots, r;\ j = r+3, r+4;\ \rho_{ij} \neq 0),$$

$$\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$$

$$(X_i, X_j) = \rho_{ij} \sum_{s=r+2k-1}^{r+2k} c_{ijs} X_s \ (i=1,\ 2,\ \cdots,\ r;\ j=r+2k-1,\ r+2k;\ \rho_{ij}\neq 0),$$

and

$$(X_i, X_j) = 0$$

$(i = r + 1,\ \cdots,\ r + 2k;\ \ j = i + 1$, when $i = r +$ odd number, and $j = i - 1$, when $i = r +$ even number).

Denoting the operators of the adjoint of G_{r+2k} by the symbols $E_1, \cdots, E_r, E_{r+1}, \cdots, E_{r+2k}$, we have for the alternant of any two of these operators the same structural constants as for the alternant of the corresponding operators of the group G_{r+2k}.

We shall assume in what follows that the group G_{r+2k} is meroëdrically isomorphic with a simple group G_r of order r having one invariant spread (not flat).

* See the preceding paper.

Denoting the operators of G_r by $Y_1, \cdots, Y_r$ and the operators of the adjoint of G_r by $A_1, \cdots, A_r$, we have

$$(Y_i, Y_j) = \sum_{s=1}^{r} c_{ijs} Y_s$$

and $\qquad (i, j = 1, 2, \cdots, r)$

$$(A_i, A_j) = \sum_{s=1}^{r} c_{ijs} A_s.$$

As I have already shown in my previous articles* the matrix $\Sigma \alpha_j A_j$, where, the α's being the parameters of the group G_r,

$$\Sigma \alpha_j A_j = \begin{pmatrix} \Sigma \alpha_j c_{j11}, & \cdots, & \Sigma \alpha_j c_{j1r} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \Sigma \alpha_j c_{jr1}, & \cdots, & \Sigma \alpha_j c_{jrr} \end{pmatrix},$$

is of nullity one. Let us now consider the equation

$$(A) \qquad \left(\sum_1^{r+2k} \alpha_i X_i, \sum_1^{r+2k} \beta_j X_j \right) = \sum_{i=1}^{r+2k} \sum_{j=1}^{r+2k} \sum_{s=r+1}^{r+2k} \rho_{ij} \alpha_i \beta_j c_{ijs} X_s.$$

For the α's assigned, and ρ_{ij} and c_{ijs} $(i = 1, \cdots, r;\ s, j = r+1, \cdots, r+2k)$ easily calculated, it is clear that equation (A) has the following $2k+1$ independent sets of β's for solutions:

$$\begin{array}{ll}
(1) & \beta_1 = \alpha_1, \quad \cdots, \quad \beta_r = \alpha_r, \quad \beta_{r+1} = \cdots = \beta_{r+2k} = 0 \\
(2) & \beta_1 = \cdots = \beta_r = 0, \quad \beta_{r+1} = 1, \quad \beta_{r+2} = \cdots = \beta_{r+2k} = 0 \\
(3) & \beta_1 = \cdots = \beta_r = \beta_{r+1} = 0, \quad \beta_{r+2} = 1, \quad \beta_{r+3} = \cdots = \beta_{r+2k} = 0 \\
 & \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
 & \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
(2k+1) & \beta_1 = \cdots = \beta_{r+2k-1} = 0, \quad \beta_{r+2k} = 1.
\end{array}$$

The matrix of the coefficients of the β's in the equations obtained from equation (A), namely,

$$\begin{pmatrix}
\Sigma \alpha_i c_{i11}, & \Sigma \alpha_i c_{i21}, & \cdots, & \Sigma \alpha_i c_{ir1}, & \cdots, & \Sigma \alpha_i c_{i,\,r+2k,\,1} \\
\Sigma \alpha_i c_{i12}, & \Sigma \alpha_i c_{i22}, & \cdots, & \Sigma \alpha_i c_{ir2}, & \cdots, & \Sigma \alpha_i c_{i,\,r+2k,\,2} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\Sigma \alpha_i c_{i1r}, & \Sigma \alpha_i c_{i2r}, & \cdots, & \Sigma \alpha_i c_{irr}, & \cdots, & \Sigma \alpha_i c_{i,\,r+2k,\,r} \\
\multicolumn{6}{l}{\Sigma \alpha_i c_{i,\,1,\,r+1}(1-\rho_{i1}), \quad \cdots, \quad \Sigma \alpha_i c_{i,\,r,\,r+1}(1-\rho_{ir}),} \\
\multicolumn{6}{r}{\cdots, \quad \Sigma \alpha_i c_{i,\,r+2k,\,r+1}(1-\rho_{i,\,r+2k})} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\multicolumn{6}{l}{\Sigma \alpha_i c_{i,\,1,\,r+2k}(1-\rho_{i1}), \quad \cdots, \quad \Sigma \alpha_i c_{i,\,r,\,r+2k}(1-\rho_{ir}),} \\
\multicolumn{6}{r}{\cdots, \quad \Sigma \alpha_i c_{i,\,r+2k,\,r+2k}(1-\rho_{i,\,r+2k})}
\end{pmatrix},$$

will *have for its nullity a number not less than* $2k+1$. It follows, therefore, since every minor of the determinant $|\Sigma \alpha_i A_i|$ is also a minor of the

* Loc. cit.

determinant of the matrix obtained from equation (A), that the nullity of the matrix $\Sigma\alpha_i E_i$ is exactly equal to $2k+1$. The nullity of the matrix $\Sigma\alpha_i E_i$ is equal to the number of independent invariants of the adjoint of G. The following *theorem* can therefore be stated:

If the adjoint of G_r which is meroëdrically isomorphic with G_{r+2k} has one invariant, then the adjoint of G_{r+2k} has $2k+1$ independent invariants, one of which is the invariant of the adjoint of G_r.

2. We shall denote the $2k+1$ invariants by $V(\alpha_1, \cdots, \alpha_r)$, $W_1(\alpha_1, \cdots, \alpha_r, \cdots, \alpha_{r+2k})$, $\cdots$, $W_{2k}(\alpha_1, \cdots, \alpha_r, \cdots, \alpha_{r+2k})$; the equations

$$\begin{aligned} V(\alpha) &= 0 \\ W_1(\alpha) &= 0 \\ \cdot \quad \cdot \quad &\cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad &\cdot \quad \cdot \quad \cdot \\ W_{2k}(\alpha) &= 0 \end{aligned}$$

will then represent $2k+1$ invariant spreads in the space of the adjoint of G_{r+2k}. Now, since the group G_{r+2k} has k invariant two-parameter subgroups, the adjoint of G_{r+2k} will leave invariant k straight lines, each one representing the corresponding invariant subgroup. We shall denote those lines, as in the previous paper, by the symbols $X_i \leftarrow\rightarrow X_j$ ($i = r+1, \cdots, r+2k$; $j = i+1$, when $i = r +$ odd number, and $j = i-1$, when $i = r +$ even number).

If the equations $W_1(\alpha) = 0, \cdots, W_{2k}(\alpha) = 0$ represent $2k$ invariant flats in the space of the adjoint of G_{r+2k}, then their common intersection (if there is any) will be an $(r-1)$-flat also invariant to the adjoint of G_{r+2k}. If that flat does not pass through any of the lines $X_i \leftarrow\rightarrow X_j$, then by Lie's theorem* we can take $\overline{X}_1, \cdots, \overline{X}_r$, $\overline{X}_{r+1} = X_{r+1}, \cdots, \overline{X}_{r+2k} = X_{r+2k}$, such linear functions of $X_1, \cdots, X_{r+2k}$ that the first $r-1$ operators form an invariant subgroup of order $r-1$, while the last $2k$ operators still generate the invariant two-parameter subgroups which we started with, i.e.,

$$(\overline{X}_i, \overline{X}_j) = \sum_{s=1}^{r-1} \bar{c}_{ijs} \overline{X}_s \qquad (i = 1, \cdots, r-1;\ j = 1, \cdots, r),$$

$$(\overline{X}_i, \overline{X}_j) = 0 \qquad (i = 1, \cdots, r-1;\ j = r+1, \cdots, r+2k),$$

$$(\overline{X}_i, \overline{X}_j) = \bar{\rho}_{ij} \sum_{r-1}^{r+2} \bar{c}_{ijs} \overline{X}_s \qquad (i = r, r+1, r+2;\ j = r+1, r+2),$$

$$(\overline{X}_i, \overline{X}_j) = \bar{\rho}_{ij} \sum_{r+3}^{r+4} \bar{c}_{ijs} \overline{X}_s \qquad (i = r, r+3, r+4;\ j = r+3, r+4),$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$(\overline{X}_i, \overline{X}_j) = \bar{\rho}_{ij} \sum_{r+2k-1}^{r+2k} c_{ijs} \overline{X}_s \qquad (i=r, r+2k-1, r+2k;\ j=r+2k-1, r+2k),$$

$$(\overline{X}_i, \overline{X}_j) = 0 \qquad (i = r+1, \cdots, r+2k;\ j = i \pm 2;\ j \neq r).$$

* Lie-Scheffers, Continuierliche Gruppen, p. 479.

If however not all flats have a common intersection, then the flat representing the common intersection of a few will enable us to form an invariant subgroup of G_{r+2k}.

Suppose now that $W_1(\alpha) = 0, \cdots, W_{2k}(\alpha) = 0$ are spreads of degrees $m_1, m_2, \cdots, m_{2k}$ respectively. Consider then the polar flats, with respect to each of the spreads

$$W_1(\alpha) = 0, \qquad \cdots, \qquad W_{2k}(\alpha) = 0,$$

of each of the lines $X_{r+1} \leftarrow\rightarrow X_{r+2}, \cdots, X_{r+2k-1} \leftarrow\rightarrow X_{r+2k}$. Taking, for instance, the polar flats of the line $X_{r+1} \leftarrow\rightarrow X_{r+2}$, with respect to

$$W_1(\alpha) = 0, \qquad \cdots, \qquad W_{2k}(\alpha) = 0,$$

we obtain $2k$ invariant r-flats in the space of the adjoint of G_{r+2k}. The same holds for all the other lines. Thus, there will be $2k^2$ invariant flats in addition to the k invariant flats formed by taking the polar flats of those lines with respect to the spread $V(\alpha) = 0$ ($V(\alpha)$ is the common invariant of the adjoints of G_r and G_{r+2k}).

If the common intersection of those $2k^2 + k$ invariant r-flats, if there is any, is an $(r - 1)$-flat, we can choose the operators of G_{r+2k} in such a way that $r - 1$ of them form an invariant subgroup of order $r - 1$. Or, more generally, if the common intersection is an $(r - i)$-flat ($1 \leqq i \leqq r - 1$), the group G_{r+2k} will *have an invariant subgroup of order* $r - i$ by properly choosing the operators.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
December 7, 1920.

THE AUTOMORPHIC TRANSFORMATION OF A BILINEAR FORM.

By J. H. M. Wedderburn.

1. **Introduction.** The problem of transforming a bilinear form into itself cogrediently was first solved by Hermite* and Cayley for symmetric and skew-symmetric forms and later by Voss† for any form. This solution has two defects; in the first place it only gives transformations whose determinant is $+1$, and secondly it becomes indeterminate for those transformations whose characteristic roots include both $+1$ and -1. These exceptional cases have been treated more or less completely by a number of authors.

The aim of this note is to present a method of obtaining a form for the automorphic transformation which displays clearly the rôle played by the exceptional cases. The parameters in this solution enter transcendentally but it is free from the first kind of exception; and in deriving from it the Hermite-Cayley form, which is rational, the analytical nature of this exceptional case is made clear.

The exponential and logarithmic functions of a matrix form the basis of the exposition and in view of this it has been thought advisable to include a short discussion of functions of a matrix in general especially as, in spite of the fact that there is little or nothing new in the results obtained, there is no place in the literature where the necessary properties are collected together.

2. **The idempotent units of a matrix.** If the elementary divisors of a matrix x are $(\lambda - g_i)^{\nu_i}$ $(i = 1, 2, \cdots, r)$, then, when the basis is properly chosen, x can be expressed‡ as the direct sum of irreducible matrices of the form

$$(1) \qquad x_i = \begin{vmatrix} g_i & 1 & & & \\ & g_i & 1 & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & g_i & 1 \\ & & & & g_i \end{vmatrix} = g_i e_i + \eta_i, \quad (i = 1, 2, \cdots, r)$$

* Hermite, "Sur la théorie des formes quadratiques ternaires indéfinies," Crelle, 47 (1854), pp. 307–312; Cayley, "A memoir on the automorphic linear transformation of a bipartite quadric," Lond. Phil. Trans., 148 (1858), pp. 39–46. For further references see Encyc. des Sci. Math., I, 2, fasc. 4, p. 489.

† Voss, "Über die cogredienten Transformationen einer bilinearen Form in sich selbst," Münch. Abh., 17 (1892), pp. 235–356.

‡ Cf. Bôcher, Higher algebra, p. 289.

where e_i and η_i are matrices of rank p_i and $p_i - 1$, respectively, which satisfy the conditions

$$(2) \quad e_i^2 = e_i, \quad \eta_i^{p_i} = 0, \quad \eta_i^{p_i-1} \neq 0, \quad e_i\eta_i = \eta_i = \eta_i e_i, \quad e_i e_j = 0 \quad (i \neq j).$$

We shall say that e_i and η_i are the idempotent and nilpotent units corresponding to the elementary divisor $(\lambda - g_i)^{p_i}$. These units are not unique when the same root occurs in several elementary divisors. For instance, if x is the matrix

$$\begin{Vmatrix} g & 0 & 0 \\ 0 & g & 1 \\ 0 & 0 & g \end{Vmatrix},$$

or, using matric units* e_{pq},

$$x = ge_{11} + g(e_{22} + e_{33}) + e_{23},$$

where $e_1 = e_{11}$, $e_2 = e_{22} + e_{33}$ and $\eta_2 = e_{23}$; then if we set

$$\begin{array}{lll} a_{11} = e_{11} - e_{13} & a_{12} = e_{12} & a_{13} = e_{13} \\ a_{21} = e_{21} - e_{23} & a_{22} = e_{22} & a_{23} = e_{23} \\ a_{31} = e_{11} - e_{13} + e_{31} - e_{33} & a_{32} = e_{32} + e_{12} & a_{33} = e_{33} + e_{13}, \end{array}$$

the a's form a set of matric units and

$$\begin{aligned} x &= g(e_{11} - e_{13}) + g(e_{22} + e_{33} + e_{13}) + e_{23} \\ &= ga_{11} + g(a_{22} + a_{33}) + a_{23}, \end{aligned}$$

so that $e_{11} - e_{13}$ and $e_{22} + e_{33} + e_{13}$ might have been chosen as idempotent units in place of e_1 and e_2.

It is shown below that, in any representation of x in the form (1), the sum e_j of all the idempotent units which belong to the same root g_j is independent of the particular representation used. It will be called the *principal* unit corresponding to g_j, its parts being called *partial* units.

It should be noticed that other normal forms are possible: for instance, in place of (1) we can by a different choice of η_i express x_i in the form

$$(3) \qquad x_i = g_i e_i + h_{i1}\eta_i + h_{i2}\eta_i^2 + \cdots + h_{i,\, p_i-1}\eta_i^{p_i-1},$$

where the h's are preassigned constants different from zero.

3. **Functions of a matrix.** Using the notation of the last paragraph, let

$$(4) \qquad x = \Sigma x_i, \qquad x_i = g_i e_i + \eta_i, \qquad (i = 1, 2, \cdots, r)$$

be an expression of x as the sum of irreducible matrices, then

$$x_i^m = (g_i e_i + \eta_i)^m = g_i^m e_i + mg_i^{m-1}\eta_i + \binom{m}{2} g_i^{m-2}\eta_i^2 + \cdots,$$

* The unit e_{pq} is a matrix for which the coefficient in the pth row and qth column is 1, while all the other coefficients are zero. The law of combination of these units is $e_{pq}e_{qr} = e_{pr}$, $e_{pq}e_{sr} = 0$ $(q \neq s)$.

the binomial expansion terminating with the $(m+1)$th term when m is less than the rank p_i of e_i, and with the p_ith term when this is not the case.

If now $f(\lambda)$ is any function expansible in a Taylor series which converges for every root of x, then $f(x)$ is reducible in the same way as x, the part corresponding to x_i being

$$(5)\quad f_i(x) = f(g_i)e_i + f'(g_i)\eta_i + f''(g_i)\frac{\eta_i^2}{2!} + \cdots + f^{(p_i-1)}(g_i)\frac{\eta_i^{p_i-1}}{(p_i-1)!}$$

or, writing (5) in full but with the subscript i omitted, we have

$$\left|\begin{array}{cccccc} f(g) & f'(g) & \dfrac{f''(g)}{2!} & \dfrac{f'''(g)}{3!} & \cdots & \dfrac{f^{(p-1)}(g)}{(p-1)!} \\ & f(g) & f'(g) & \dfrac{f''(g)}{2!} & \cdots & \dfrac{f^{(p-2)}(g)}{(p-2)!} \\ & & f(g) & f'(g) & \cdots & \dfrac{f^{(p-3)}(g)}{(p-3)!} \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \cdot \\ & & & & & f(g) \end{array}\right|$$

where there are p_i rows and columns and every term to the left of the main diagonal is zero while, in the main diagonal itself and on its right, the terms in the first row are repeated in the succeeding rows, all terms lying on a parallel to the main diagonal being the same. An important particular case is $f(\lambda) \equiv \exp\lambda = \epsilon^\lambda$ for which

$$(6)\quad f_i(x) = \epsilon^{g_i}\left(e_i + \eta_i + \frac{\eta_i^2}{2!} + \cdots + \frac{\eta_i^{p_i-1}}{(p_i-1)!}\right) = e_i + x_i + \frac{x_i^2}{2!} + \frac{x_i^3}{3!} + \cdots.$$

Suppose now that $y = \Sigma y_i$ is a matrix whose reduced parts have the form

$$y_i = g_i^{(0)}e_i + g_1^{(1)}\eta_i + \cdots + g_i^{(p_i-1)}\frac{\eta_i^{p_i-1}}{(p_i-1)!},$$

where e_i and η_i are the same as in (4), i.e., belong to x, and the g's are any set of constants. Then by the extension of Lagrange's interpolation formula (see § 4 below), there is a polynomial $f(\lambda)$ for which

$$f(g_i) = g_i^{(0)},\quad f'(g_i) = g_i^{(1)},\quad \cdots,\quad f^{(p_i-1)}(g_i) = g_i^{(p_i-1)},\quad (i = 1, 2, \cdots)$$

so that we may set $y = f(x)$. In particular if we set

$$g_i^{(0)} = \log g_i + 2k_i\pi\sqrt{-1} \equiv f(g_i)$$
$$g_1^{(j)} = \frac{d^j \log g_i}{dg_i^j} = f^{(j)}(g_i)$$

and also put $F(\lambda) = \epsilon^{f(\lambda)}$, then

$$F_i(x) = \epsilon^{f(g_i)}[e_i + f'(g_i)\eta_i + (f'^2(g_i) + f''(g_i))\eta_i^2 + \cdots]$$
$$= g_i e_i + \eta_i = x_i,$$

since the coefficients of the various powers of η_i are formally the successive derivatives of $\epsilon^{\log g_i}$. We have therefore $F(x) = x$, so that we may set $f(x) = \log x$. The logarithmic function so defined is indeterminate to an additive term of the form $2\pi i \Sigma k_i e_i$ where e_i $(i = 1, 2, \cdots)$ is any set of idempotent units belonging to x and the k's are integers. It is fairly obvious that any function which possesses the necessary derivatives may be extended to the case of a matric variable in a similar fashion.

Considerable care must be exercised in using the logarithmic function. For instance, if x and y are commutative, $\log x$ and $\log y$ will also be commutative if the same determination of $\log g_i$ is used with all the partial units depending on the root g_i; for the principal units of commutative matrices are commutative. If however this precaution is not taken, it is no longer true that $\log x$ and $\log y$ are necessarily commutative. For instance, if x is the matrix already used as an illustration in § 2 and $\operatorname{Log} g$ is a particular determination of $\log g$, then

$$z_1 = (\operatorname{Log} g + 2\pi i)e_{11} + \operatorname{Log} g\,(e_{22} + e_{33}) + \frac{1}{g}e_{23}$$

and

$$z_2 = \operatorname{Log} g\,(e_{11} - e_{13}) + \operatorname{Log} g\,(e_{22} + e_{33} + e_{13}) + \frac{1}{g}e_{23} = z_1 - 2\pi i e_{11}$$

are two determinations of $\log x$ which are not commutative. In this paper we shall only require logarithms in which the condition given above is satisfied and to indicate this we shall write $\operatorname{Log} x$ in place of $\log x$, so that $\operatorname{Log} x$ is determinate to an additive term of the form $\Sigma 2\pi k i e_j$ where the e_j are the *principal* units of x. The principal idempotent units of $\operatorname{Log} x$ are then the same as those of x while, as in (5), its principal nilpotent units are scalar polynomials of the corresponding principal nilpotent units of x.

The same difficulties arise, of course, with any multiple-valued function.

4. **The interpolation formula.** As we shall have need of it later, we shall now develop the generalization* of the Lagrange interpolation formula referred to in the previous section. Let

$$(7)\qquad \varphi(x) = (x - g_1)^{p_1}(x - g_2)^{p_2} \cdots (x - g_r)^{p_r}, \quad \left(\sum_1^r p_i = n, \ \ r > 1\right),$$

be the reduced equation of x, the roots g being all distinct. If we set

* Cf. Encyc. des Sci. Math., I, 2, fasc. 1, p. 61.

$$P_i(x) = \prod_{j \neq i} \left(\frac{x - g_j}{g_i - g_j}\right)^{p_j},$$

we can determine two polynomials $Q_i(x)$ and $D_i(x)$ of degree $p_i - 1$ and $n - p_i - 1$ respectively such that

$$P_i(x)Q_i(x) + (x - g_i)^{p_i}D_i(x) \equiv 1.$$

Setting

(9) $$R_i(x) = P_i(x)Q_i(x),$$

$1 - \Sigma_i R_i(x)$ is divisible by $\varphi(x)$ and, being of degree $n - 1$, at most is therefore zero; hence

(10) $$\sum_i R_i(x) \equiv 1.$$

If $h(x)$ is any polynomial in x with scalar coefficients, then

$$\begin{aligned} h(x) &= \Sigma h(x)R_i(x) \\ &= \Sigma\left[h(g_i) + h'(g_i)(x - g_i) + \cdots + \frac{h^{(p_i-1)}(g_i)}{(p_i - 1)!}(x - g_i)^{p_i-1}\right]R_i(x) \\ &\qquad + \Sigma C_i(x)(x - g_i)^{p_i}R_i(x), \end{aligned}$$

where C_i is a polynomial, being in fact the coefficient of $(x - g_i)^{p_i}$ in the remainder when $h(x)$ is expanded in a Taylor series. Now it follows from the definition of R_i that $(x - g_i)^{p_i}R_i(x)$ is divisible by $\varphi(x)$, hence, setting $R_{ij}(x) = (x - g_i)^j R_i(x)$,

(11) $$h(x) \equiv \sum_{i=1}^{r} \sum_{j=0}^{p_i-1} \frac{h^j(g_i)}{j!} R_{ij}(x) \quad (\text{mod } \varphi(x)).$$

If h is of lower degree than φ, this congruence is an algebraic identity, and therefore gives the form of a polynomial which, along with its derivatives up to the $(p_i - 1)$th order, has arbitrarily assigned values for $x = g_i$ $(i = 1, 2, \cdots, r)$.

When x is a matrix and $\varphi(x) = 0$ is its reduced equation, then (11) is again an identity in the *coefficients* of x and gives the form of any scalar polynomial in x. Since $R_i^2 = R_i$, $R_iR_j = 0$, $(i \neq j)$ and

$$R_{ij} = (x - g_i)^j R_i(x),$$

it is easily seen that R_i is what we have already called the principal idempotent unit belonging to g_i, and R_{i1} is the sum of the units η which belong to this root, i.e., it is the corresponding principal nilpotent unit; we may also notice here that $R_{i1}{}^j = R_{ij}$.

The principal units are therefore scalar polynomials in x, a result which is of some importance in the sequel.

The above argument requires some modification when h is not a

polynomial but, in view of what has already been said in the previous paragraph, it is not necessary to discuss the matter here.

Functions of two or more commutative matrices can be treated in a similar fashion.* Let x and y be two commutative matrices whose roots are $g_1, g_2, \cdots$ and $h_1, h_2, \cdots$, respectively, and as above let $R_i(x)$ and $R_i(y)$ $(i = 1, 2, \cdots)$ denote the principal units of these matrices. Then, if we set

$$S_{ij} = R_i(x)R_j(y),$$

those S_{ij} which are not zero are linearly independent. For, if $\Sigma \xi_{ij} S_{ij} = 0$, then

$$0 = R_p(x)\Sigma\xi_{ij}S_{ij}\cdot R_q(y) = \xi_{pq}S_{pq},$$

so that $\xi_{pq} = 0$ unless $S_{pq} = 0$.

From the definition of S_{ij} it follows that $S_{ij}S_{pq} = 0$ if $i \neq p$ or $j \neq q$, also $S_{ij}{}^2 = S_{ij}$ and $\Sigma_{ij}S_{ij} = 1$; hence

$$x = \sum_{i,j}[g_i + (x - g_i)]S_{ij}, \qquad y = \sum_{i,j}[h_j + (y - h_j)]S_{ij},$$

where $(x - g_i)S_{ij}$ and $(y - h_j)S_{ij}$ are commutative nilpotent matrices.

If $\psi(x, y)$ is any scalar polynomial in x and y, we may now set

$$\begin{aligned}\psi(x, y) &= \sum_{r,s}\psi_{rs}{}^{ij}(x - g_i)^r(y - h_j)^s \\ &= \sum_{i,j}\Big[\psi(g_i, h_j)S_{ij} + \sum_{r,s}\psi_{rs}{}^{ij}(x - g_i)^r(y - h_j)^s S_{ij}\Big],\end{aligned}$$

where in the second summation r and s are not both zero; or, if we let $(x - g_i)^r(y - h_j)^s S_{ij} = S_{ij}{}^{rs}$, then

$$\begin{aligned}\psi(x, y) &= \sum_{i,j}\psi(g_i, h_j)S_{ij} + \sum_{i,j}\sum_{r,s}\psi_{rs}{}^{ij}S_{ij}{}^{rs} \\ &= z + w,\end{aligned}$$

say, where w is nilpotent, being the sum of a number of commutative nilpotent matrices. Now if $\varphi(z) = 0$ is the reduced equation of a matrix z, and w is a nilpotent matrix commutative with z for which $w^s = 0$, then, if $F(z) = \varphi^s(z)$, we have

$$F(z + w) = F(z) + F'(z)w + \cdots + \frac{F^{(s-1)}(z)}{(s-1)!}w^{s-1} = 0,$$

since the first s derivatives of $F(z)$ are divisible by $\varphi(z)$ and therefore vanish. It follows that the characteristic of the reduced equation of $z + w$ is a factor of a power of that of z, and vice versa; hence the roots of z and $z + w$ are the same. We can say, therefore, that *if $R_i(x)$ and $R_j(y)$ $(i, j = 1, 2, \cdots)$ are the principal idempotent units of two commuta-*

* Cf. Frobenius, "Über vertauschbare Matrizen," Berl. Sitzb. (1896), pp. 601–614.

tive matrices x and y, and $S_{ij} = R_i(x)R_j(y)$; and if g_i and h_j are the corresponding roots of x and y; then the roots of any scalar function $\psi(x, y)$ of x and y are $\psi(g_i, h_j)$ where i and j take only those values for which $S_{ij} \neq 0$.

The extension to functions of several commutative matrices is obvious.

5. **The automorphic transformation of a matrix.** If y is a non-singular matrix, the problem of transforming it into itself is equivalent to finding all the matric solutions of the equation*

$$x'yx = y. \tag{12}$$

When solved for x', this equation gives

$$x' = yx^{-1}y^{-1}, \tag{13}$$

from which it follows immediately that the identical equation of x has reciprocal roots and that, if g is any root other than ± 1, the elementary divisors corresponding to g and $1/g$ occur in pairs† with the same exponents. It follows also from (13) that

$$x' = yx^{-1}y^{-1} = y'x^{-1}y'^{-1}$$

so that x is commutative with $y^{-1}y'$.

If $h(\lambda)$ is a scalar polynomial in λ, then from (13) $h(x') = yh(x^{-1})y^{-1}$, and therefore, in particular, the principal unit of x' corresponding to a root $g_i \neq \pm 1$ is the transform of the principal unit‡ of x corresponding to $1/g_i$. If we denote the principal units belonging to g_i ($g_i \neq \pm 1$) and $1/g_i$ by e_i and e_{-i}, respectively, we have therefore

$$e_i' = ye_{-i}y^{-1}; \tag{14}$$

and similarly, if e_1 and e_{-1} belong to the roots ± 1 when these roots are present, we have

$$e_1' = ye_1y^{-1}, \qquad e_{-1}' = ye_{-1}y^{-1}. \tag{15}$$

If now we set

$$x = \epsilon^z, \qquad z = \operatorname{Log} x, \tag{16}$$

then from (12)

$$1 = x'yxy^{-1} = \epsilon^{z'}\epsilon^{yzy^{-1}} = \epsilon^{z'+yzy^{-1}}. \tag{17}$$

Here $z' + yzy^{-1}$ has the same principal idempotent units as x' and z', and hence it has the form $\Sigma(\gamma_i e_i' + \bar{\eta}_i')$ where $\bar{\eta}_i'$ is a scalar polynomial in η_i'; hence (17) is equivalent to

$$1 = \epsilon^{z'+yzy^{-1}} = \Sigma\epsilon^{\gamma_i}\left(e_i' + \bar{\eta}_i' + \frac{\bar{\eta}_i'^2}{2!} + \cdots\right),$$

* Here x' denotes, as usual, the transverse or conjugate of x.

† Cf. Kronecker, Crelle, 68 (1868), p. 273.

‡ Cf. Taber, "On the automorphic linear transformation of an alternate bilinear form," Math. Ann., 46 (1895), p. 568. The principal unit of x belonging to g is the same as the principal unit of $1/x$ belonging to $1/g$.

whence $\overline{\eta}_i' = 0$ and* $\gamma_i = 2k_i\pi\iota$. We can therefore set

$$z' + yzy^{-1} = 2\pi\iota\Sigma k_i e_i'. \tag{18}$$

Since x, and therefore also z, is commutative with $y^{-1}y'$, we have

$$y'^{-1}z'y' + z = 2\pi\iota\Sigma k_i y'^{-1}e_i'y'$$

or, forming the transverse of each side and using (14) and (15),

$$z' + yzy^{-1} = 2\pi\iota\Sigma k_i y e_i y^{-1} = 2\pi\iota(k_1e_1' + k_{-1}e_{-1}' + \Sigma k_i e_i') \quad (i \neq \pm 1),$$

and, comparing this with (18), we have

$$k_i = k_{-i} \qquad (i \neq \pm 1). \tag{19}$$

We can now simplify (18) as follows. Set

$$z_1 = z - 2\pi\iota(\Sigma' k_i e_i + \lambda_1 e_1 + \lambda_{-1}e_{-1}),$$

where in the summation sign the prime indicates that the roots $g_i \neq \pm 1$ are arranged in pairs g_i and $1/g_i$ and only the first of each pair is taken in forming the sum. Inserting z_1 in place of z in (18) we have from (14) and (15)

$$\begin{aligned} z_1' + yz_1y^{-1} = z' + yzy^{-1} &- 2\pi\iota(\Sigma' k_i e_i' + \lambda_1 e_1' + \lambda_{-1}e_{-1}' \\ &\qquad + \Sigma' k_i e_{-i}' + \lambda_1 e_1' + \lambda_{-1}e_{-1}') \\ = 2\pi\iota[(k_1 &- 2\lambda_1)e_1' + (k_{-1} - 2\lambda_{-1})e_{-1}'], \end{aligned}$$

where by a proper choice of λ_1 and λ_{-1} the coefficients of e_1' and e_{-1}' may be made equal to 0 or 1. Now evidently $\epsilon^{z_1} = \epsilon^z$; it follows that there is no lack of generality in writing in place of (18)

$$z' + yzy^{-1} = 2\pi\iota\zeta', \tag{20}$$

where

$$\zeta = \theta_1 e_1 + \theta_2 e_{-1} = \zeta_1 + \zeta_2 \qquad (\theta_1, \theta_2 = 0 \text{ or } 1). \tag{21}$$

Writing now

$$w = z + \pi\iota\zeta, \tag{22}$$

equation (20) becomes

$$w' + ywy^{-1} = 0, \tag{23}$$

which may also be written

$$y'(wy^{-1})' + y(wy^{-1}) = 0,$$

or, if $u = wy^{-1}$,

$$y'u' + yu = 0, \tag{23'}$$

which is equivalent to the equation given by Cayley.† The solution of

* Here $\iota = \sqrt{-1}$.

† Cayley, l.c., p. 44. Cayley's solution is incomplete as he omits to impose the necessary conditions on the skew-symmetric matrix which enters into his result; and this leads him to draw erroneous conclusions.

(23) which is given in the next section is practically that given by Voss.*

6. **The equation** $w' + ywy^{-1} = 0$. We shall consider in place of (23) the more general equation

$$w' = \delta ywy^{-1} \qquad (\delta = \pm 1). \tag{24}$$

Forming the transverse of each side, we get $w = \delta y'^{-1}w'y'$ or $w' = \delta y'wy'^{-1}$, whence

$$wy^{-1}y' = y^{-1}y'w \qquad \text{or} \qquad y'wy'^{-1} = ywy^{-1}, \tag{25}$$

i.e., w is commutative with $y^{-1}y'$. Now from (24) we have $w = \delta y^{-1}w'y$, so that $2w = w + \delta y^{-1}w'y$. But if v is any matrix commutative with $y^{-1}y'$, then

$$w = v + \delta y^{-1}v'y \tag{26}$$

is a solution of (24) as, on substituting this value for w, we get

$$w' - \delta ywy^{-1} = v' + \delta y'vy'^{-1} - \delta yvy^{-1} - v' = 0,$$

since $y'vy'^{-1} = yvy^{-1}$. The most general solution of (23) is therefore obtained by setting

$$w = v - y^{-1}v'y, \qquad vy^{-1}y' = y^{-1}y'v. \tag{27}$$

It should be noted, however, that two different values of v may lead to the same value of w.

When $\delta = -1$, we have relations among the roots and idempotent units of w which are the logarithmic counterpart of those already given for x. For, since

$$|\lambda - w| = |\lambda - w'| = |\lambda + ywy^{-1}| = |\lambda + w|, \tag{28}$$

the non-zero roots of w occur in pairs of opposite sign and with equal exponents in the elementary divisors. We can show exactly as in § 5 that if e_i is the principal unit corresponding to a root g_i ($g_i \neq 0$) and e_{-i} the principal unit belonging to $-g_i$, then

$$e_i' = ye_{-i}y^{-1}, \qquad e_{-i}' = ye_iy^{-1}, \tag{29}$$

and if e_0 is the principal unit belonging to the root 0, if present, then

$$e_0' = ye_0y^{-1}. \tag{30}$$

Since $(w')^r = (-1)^r yw^ry^{-1}$, the reduced equation of w has the form $w^m\psi(w^2) = 0$; hence e_0 is a polynomial in w^2, which gives an independent proof of (30), since $(w^2)' = yw^2y^{-1}$.

The form of w given in (22) can be still further simplified by means of these relations. In (21) the term ζ_1 is the sum of partial units of x

* Voss, l.c., p. 330.

coming from roots equal to unity and they therefore correspond to roots of the form $2\pi k\iota$ of z, where k is integral, and hence to roots $\pi(2k+1)\iota$ of w. Let a_1 be the principal unit of w corresponding to this root and a_2 that belonging to its negative $-\pi(2k+1)\iota$ so that by (29)

$$a_1' = ya_2y^{-1};$$

and let ζ_{11} be that part of ζ_1 which is a partial unit of a_1wa_1 so that $\zeta_{11} = a_1\zeta_1 = \zeta_1 a_1$; then, if $\zeta_{22} = \zeta_1 a_2 \zeta_1$, we have

$$\zeta_{11}' = \zeta_1' a_1' \zeta_1' = y\zeta_1 a_2 \zeta_1 y^{-1} = y\zeta_{22}y^{-1};$$

and similarly

$$\zeta_{22}' = y\zeta_{11}y^{-1}.$$

But $\zeta_{11} = \zeta\zeta_{11}\zeta$; therefore $\zeta_{11}' = y\zeta\zeta_{22}\zeta y^{-1}$, so that $\zeta\zeta_{22}\zeta = \zeta_{22}$ and ζ_{22} is therefore also a partial unit along with ζ_{11}; the rank of $\zeta_1 - \zeta_{11} - \zeta_{22}$ is less than that of ζ_1.

If now we put

$$\bar{z} = z - 2\pi\iota\,\zeta_{11},$$

we have as before

$$\begin{aligned}\bar{z}' + y\bar{z}y^{-1} &= 2\pi\iota(\zeta_1' + \zeta_2') - 2\pi\iota\zeta_{11}' - 2\pi\iota y\zeta_{11}y^{-1}\\ &= 2\pi\iota(\zeta_1' - \zeta_{11}' - \zeta_{22}') + 2\pi\iota\zeta_2'.\end{aligned}$$

This transformation therefore replaces ζ_1 by a new ζ with lower rank and at the same time does not alter x. By repeating this process we can reduce the rank to zero which means that we can assume $\zeta_1 = 0$ without loss of generality.

In the same way ζ_{-1} corresponds to roots $(2k+1)\pi\iota$ of z and therefore to roots $(2k+2)\pi\iota$ of w. If $k \neq -1$, the rank of ζ_{-1} can be reduced as above so that it is only necessary to take account of zero roots of w in considering the form of ζ_{-1}.

7. **The determination of z.** The results of the preceding paragraph may be summarized by saying that every value of x in (12) can be obtained by putting

$$z = w + \theta\pi\iota\zeta, \qquad (\theta = 0,\, 1) \tag{31}$$

where w is any solution of (23) and ζ is an idempotent matrix corresponding to a zero root of w which satisfies the equation

$$\zeta' = y\zeta y^{-1}. \tag{32}$$

In order to complete the determination of z it is therefore necessary to show how ζ is to be determined. The principal idempotent unit, e_0, belonging to the zero root* of w is of course one possible value; the only

* When the order, n, of y is odd, there must evidently be an odd number of such roots; while if n is even, there will be an even number or none.

difficulty is then to ascertain when there will exist partial units of e_0 which satisfy (32).

We shall first separate off the part of w depending on e_0 by writing

$$w = (1 - e_0)w + e_0 w = w_1 + w_0,$$

where $w_1 w_0 = 0 = w_0 w_1$. The zero roots of w_1 correspond to simple elementary divisors, and w_0 has only zero roots; both are solutions of (23).

Suppose now that e_0 can be expressed as the sum of partial units of x, say

$$e_0 = e_1 + e_2 + \cdots + e_p,$$

such that $e_i e_j = 0$ $(i \neq j)$ and $e_i' = y e_i y^{-1}$; each e_i is then a possible determination of ζ. This being so, the matrix

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_p e_p,$$

where the α's are scalars, is a solution of

$$(33) \qquad a' = yay^{-1},$$

$e_1, e_2, \cdots$ being the principal units corresponding to the roots $\alpha_1, \alpha_2, \cdots, \alpha_p$. Conversely, if a is any solution of (33), $e_0 a e_0$ is also a solution and if $e_1, e_2, \cdots, e_s$ are its prinicipal units, they are solutions of (32) and are therefore available as values of ζ. Also if ζ is a sum of any or all of these e's, then $w_0 = (1 - \zeta)w_0 + \zeta w_0$, and $(1 - \zeta)w_0$ and ζw_0 are solutions of (23) all of whose roots are zero. Further, if e is any idempotent unit which satisfies (33), ewe is a solution of (23). We therefore conclude that *every x which transforms a non-singular matrix y into itself cogrediently is of the form ϵ^z where z is determined as follows: take any solution w_1 of (23) and any solution a of (33) and let ζ be a principal unit of the latter, then*

$$z = (1 - \zeta)w_1(1 - \zeta) + \zeta w_1 \zeta + \theta\pi\iota\zeta = w + \theta\pi\iota\zeta, \quad (\theta = 0, 1).$$

The determinant of x is ± 1 according as the rank of ζ is odd or even. Here w may be any solution of (23) and is therefore a continuous function of a certain number of parameters; hence x is also a continuous function of these parameters but involves at least one other parameter θ in which it is not continuous since the part of $\log x$ which depends on θ vanishes except for $\theta = 1$.

We may also notice here that we can set

$$x = \epsilon^w (1 - 2\theta\zeta),$$

where $(1 - 2\theta\zeta)^{-1} = (1 - 2\theta\zeta)$; and if $w = (1 - e_0)w(1 - e_0) + e_0 w e_0 = w_1 + w_0$, e_0 being the principal unit of w corresponding to its zero root,

then w_0 is nilpotent and, if

(34) $$\gamma = w_0 + \frac{w_0^2}{2!} + \cdots = e^{w_0} - 1,$$

then

(35) $$x = \epsilon^{w_1}(1 + \gamma)(1 - 2\theta\zeta), \qquad (\theta = 0, 1).$$

Here w_1 is any solution of (23) in which the zero roots have simple elementary divisors, e_0 is the principal idempotent unit corresponding to the zero root* of w_1, and γ and ζ are matrices which are respectively nilpotent and idempotent and are both solutions of

$$\varphi = e_0\varphi e_0, \qquad \varphi' = y\varphi y^{-1}.$$

8. **Rational parameters.** The parameters involved in the form of x given in the preceding paragraph enter transcendentally. If however we set

(36) $$t = \tanh\frac{z}{2} = \frac{\epsilon^z - 1}{\epsilon^z + 1} = \frac{x - 1}{x + 1},$$

then $\epsilon^z = (t - 1)/(t + 1)$ or

(37) $$x = \frac{t - 1}{t + 1},$$

also

$$yty^{-1} = \frac{yxy^{-1} - 1}{yxy^{-1} + 1} = \frac{x'^{-1} - 1}{x'^{-1} + 1} = \frac{1 - x'}{1 + x'} = -t',$$

so that t is a solution of (23); and if the coefficients of t are taken as parameters in so far as they are independent, (37) expresses x rationally in terms of these parameters. If, however, $|x + 1| = 0$, t becomes infinite, so that this form cannot give any solution which has roots equal to -1, at least directly. The difficulty arises from the fact that $\tanh(\theta/2) \to \infty$ as $\theta \to \pi\iota$, but, since $(t - 1)/(t + 1) \equiv \epsilon^z$ for all values of t which do not possess an infinite root, i.e., a root corresponding to a root $(2k + 1)\pi\iota$ of z, then x will be a solution of (12) so long as the coefficients of z are continuous functions of the parameters involved and the limiting value of x is finite and determinate. Now z is a continuous function of the parameters involved in v in equation (26) but is in general discontinuous in the coefficients of ζ; and moreover a ζ term is present only when x has a root -1. Hence if z is a solution of (17) which has no root equal to an odd multiple of $\pi\iota$, then t is finite and the expression for x in (37) remains finite even if t becomes infinite so long as z is finite and has no ζ term.

9. **Automorphic transformation of symmetric and skew-symmetric matrices: orthogonal matrices.** If y is symmetric or skew-symmetric, the matrix v

* If w_1 has no zero root, γ, θ and ζ are equal to 0.

occurring in the solution of (24) is entirely arbitrary since $y^{-1}y' = \pm 1$. Hence from (26)

$$w = v + \delta y^{-1}v'y = (vy^{-1} + \delta y^{-1}v')y = uy,$$

where $u = vy^{-1} + \delta y^{-1}v'$. Taking $\delta = -1$, u is skew-symmetric if y is symmetric and vice versa, and as any skew-symmetric (symmetric) matrix can be put in this form, u may be taken to be an arbitrary skew-symmetric (symmetric) matrix. Similarly if $\delta = +1$, the value of a in (33) becomes $a = by$, where b is an arbitrary symmetric (skew-symmetric) matrix.

If $y = 1$, then w is skew-symmetric and a symmetric; hence every orthogonal matrix has the form

$$x = \epsilon^{w}(1 - 2\theta\zeta), \qquad (\theta = 0, 1)$$

where ζ is a symmetric idempotent matrix (which may be zero) and w is a skew-symmetric matrix commutative with ζ. The known theorems regarding the roots of real orthogonal matrices are readily derived from this form.

A DIRECT DETERMINATION OF THE MINIMUM AREA BETWEEN A CURVE AND ITS CAUSTIC.

By Otto Dunkel.

If descending parallel rays of light lying in a plane fall upon a curve in that plane, a caustic* will be produced by the portions of the curve whose concavity is toward the light. The remaining portions produce no actual caustic but a virtual caustic, which becomes an actual caustic by reversing the direction of the rays. Both the actual and the virtual caustic, if any, appear in the analytical treatment. Given two points, the origin and the point $P_2(x_2, y_2)$ in the first quadrant, and a curve joining these two points with given inclinations at the points τ_1, τ_2 such that $-\pi/2 < \tau_1 < \tau_2 < \pi/2$, the area S enclosed by the curve, its caustic and the reflected rays at the two points will be considered, and the form of the curve will be determined which makes this area a minimum. Several cases of end conditions will be examined. This problem is easily treated by the methods of the Calculus of Variations,† but a more elementary method will be used here which seems to be better adapted to this special form of minimum problem, as it yields the results more directly and rapidly. The method applies in precisely the same manner to the problem of the minimum area between a curve and its evolute. It will be seen that this method is quite analogous to the high-school algebra method of solving problems of maxima and minima of quadratic functions by the device of completing the square of the quadratic function.

The Determination of the Minimizing Curve and the Area. If R is the radius of curvature of the curve at a point P at which the inclination is τ and K is the point of contact of the reflected ray with the caustic, then

$$\delta = \frac{R \cos \tau}{2} = \frac{1}{2}\frac{ds}{d\tau}\cos \tau = \frac{1}{2}\frac{dx}{d\tau}, \tag{1}$$

where $\delta = PK$ is positive when the concavity of the curve is upward and negative when downward.‡ The element of area between two reflected

* Dunkel, "Note on caustics," The American Mathematical Monthly, vol. XXVII, 1920, no. 5, page 225.

† Dunkel, "The curve which with its caustic encloses the minimum area," Washington University Studies, Scientific Series, vol. VIII, No. 2, pp. 183–194, Jan. 1921.

‡ If P and P' are neighboring points on the curve, let the normals at these two points meet in C and the reflected rays in Q. Let the perpendicular to $P'C$ at its middle point meet PC in M'; then P, P', Q and M' lie upon a circle, since $\angle PQP' = 2\angle PCP' = \angle PM'P'$. When P' approaches P, the limiting position of M' is M, the middle point of the radius of curvature R at

rays is $\delta \cos \tau\, ds/2$, which becomes $\delta^2 d\tau$ by use of (1). The integral to be made a minimum is then

$$S = \int_{\tau_1}^{\tau_2} \delta^2 d\tau, \tag{2}$$

with the two auxiliary conditions obtained from (1),

$$x_2 = 2\int_{\tau_1}^{\tau_2} \delta d\tau, \qquad y_2 = 2\int_{\tau_1}^{\tau_2} \delta \tan \tau\, d\tau. \tag{3}$$

It will be assumed that the curves considered are such that δ is a continuous function of τ. Let A and B denote two constants which will be determined later, then, after multiplying the first equation in (3) by $B/2$ and the second by $A/2$, the integral (2) may be written

$$S = \int_{\tau_1}^{\tau_2} [\delta^2 - (A \tan \tau + B)\delta] d\tau + \frac{1}{2}(Ay_2 + Bx_2).$$

This suggests the transformation to the form

$$S = \int_{\tau_1}^{\tau_2} \left[\delta - \left(\frac{A \tan \tau + B}{2}\right)\right]^2 d\tau - \int_{\tau_1}^{\tau_2} \left(\frac{A \tan \tau + B}{2}\right)^2 d\tau + \frac{1}{2}(Ay_2 + Bx_2). \tag{4}$$

Hence the minimum value of S will be given by

$$\delta = \frac{1}{2}(A \tan \tau + B), \tag{5}$$

provided that the constants A and B can be chosen uniquely so that δ satisfies the conditions (3). If this can be done, the equality (4) gives the expression for the minimum area

$$S = \frac{1}{4}(Ay_2 + Bx_2). \tag{6}$$

The constants A and B are to be determined from the equations resulting from (3),

$$\begin{aligned} x_2 &= A\int_{\tau_1}^{\tau_2} \tan \tau\, d\tau + B\int_{\tau_1}^{\tau_2} d\tau, \\ y_2 &= A\int_{\tau_1}^{\tau_2} \tan^2 \tau\, d\tau + B\int_{\tau_1}^{\tau_2} \tan \tau\, d\tau; \end{aligned} \tag{7}$$

and there exists a unique solution of these equations if their determinant

P, the limit circle has PM as a diameter and it cuts the reflected ray PQ in K, a point on the caustic. Thus $PK = \delta = R \cos \tau/2$, where $\tau = \angle MPK$. This result may also be obtained from the general formula given in The American Mathematical Monthly, l. c. Another derivation is given in the Washington University Studies, l. c.

D_2 is not zero. But this determinant is the negative of the discriminant of the quadratic form in A and B, regarded as variables,

$$A^2\int_{\tau_1}^{\tau_2}\tan^2\tau\,d\tau + 2AB\int_{\tau_1}^{\tau_2}\tan\tau\,d\tau + B^2\int_{\tau_1}^{\tau_2}d\tau = \int_{\tau_1}^{\tau_2}(A\tan\tau + B)^2d\tau.$$

Since this form is never negative and vanishes only when *both* A and B are zero, and the coefficients of the squared terms are positive, its discriminant must be greater than zero, and hence D_2, the determinant of the equations (7), must be less than zero for $\tau_2 \neq \tau_1$. For $\tau_2 = \tau_1$ it is clear that $D_2 = 0$. The constants A and B can, therefore, be determined uniquely, and hence the value of δ in (5) gives the minimum area. It may be observed that if η indicates the variation from δ as given in (5), then equation (4) may be written

$$(4') \qquad \Delta S = \int_{\tau_1}^{\tau_2}\eta^2\,d\tau,$$

where ΔS denotes the increment of the area due to the variation η. The parametric equations of the curve are obtained by integration from equations similar to (7) in which x_2, y_2, τ_2 are replaced by x, y, τ, respectively, and it will be found that

$$(8) \qquad \begin{aligned} x &= A\log\left(\frac{\sec\tau}{\sec\tau_1}\right) + B(\tau - \tau_1), \\ y &= A\left[\tan\tau - \tan\tau_1 - (\tau - \tau_1)\right] + B\log\left(\frac{\sec\tau}{\sec\tau_1}\right). \end{aligned}$$

From (1) and (5) follow the equations

$$(8') \qquad \frac{dx}{d\tau} = A\tan\tau + B, \qquad \frac{dy}{d\tau} = (A\tan\tau + B)\tan\tau,$$
$$R = (A\tan\tau + B)\sec\tau,$$

which are useful in the study of the appearance of the curve. If $A \neq 0$, they show that the curve has a cusp at the point for which $\tan\tau_0 = -B/A$. If τ_0 lies between τ_1 and τ_2, then, since δ changes sign, there is a virtual caustic given by the part of the curve for which the inclination is greater than τ_0. A discussion of the properties of the curve and of its caustic has been given in another paper,* and it is there shown that the caustic has, in general, two cusps with tangents parallel to that of the cusp of the original curve.

A Cusp at One End Point. If the minimizing curve is such that $\delta = 0$ at an end point, there is a cusp at that end and the curve is concave up from the initial point to the other end. For convenience it will be assumed

* Washington University Studies, l.c.

that the cusp is at P_2. Let τ_0 denote the inclination at this point, A_0, B_0, and δ_0 indicate the determinations of A, B, and δ for this case, so that

$$\delta_0 = \frac{1}{2}(A_0 \tan \tau + B_0).$$

If we consider any other curve passing through the origin with the inclination τ_1, and through P_2 with the inclination $\tau_2 > \tau_1$ and such that its δ does not change sign from τ_1 to τ_2, then the area given by this curve is greater than the area given by the curve δ_0. If $\tau_2 = \tau_0$, the truth of the statement follows from the previous work, so we may assume now that $\tau_2 \neq \tau_0$. For the curve δ the two equations (3) must be satisfied, while for the minimizing curve δ_0 similar equations must be written in which τ_2, δ are replaced by τ_0, δ_0. From these four equations follow the pair of equations

$$(9') \qquad \int_{\tau_1}^{\tau_2} \delta d\tau = \int_{\tau_1}^{\tau_0} \delta_0 d\tau, \qquad \int_{\tau_1}^{\tau_2} \delta \tan \tau \, d\tau = \int_{\tau_1}^{\tau_0} \delta_0 \tan \tau \, d\tau.$$

Multiplying the first equation by $B_0/2$ and the second by $A_0/2$ and adding the corresponding sides, it will be found that

$$(9) \qquad \int_{\tau_1}^{\tau_2} \delta\delta_0 \, d\tau = \int_{\tau_1}^{\tau_0} \delta_0^2 \, d\tau.$$

Comparing the two areas, we have

$$(10) \qquad \begin{aligned} \Delta S = \int_{\tau_1}^{\tau_2} \delta^2 \, d\tau - \int_{\tau_1}^{\tau_0} \delta_0^2 \, d\tau &= \int_{\tau_1}^{\tau_2} (\delta - \delta_0)^2 d\tau + 2\int_{\tau_1}^{\tau_2} \delta\delta_0 \, d\tau \\ &\quad - 2\int_{\tau_1}^{\tau_0} \delta_0^2 \, d\tau - \int_{\tau_0}^{\tau_2} \delta_0^2 \, d\tau \\ &= \int_{\tau_1}^{\tau_2} (\delta - \delta_0)^2 d\tau + \int_{\tau_2}^{\tau_0} \delta_0^2 \, d\tau, \end{aligned}$$

in which the last line follows by use of (9). Hence if $\tau_2 \leqq \tau_0$, $\Delta S > 0$ and the theorem is true in this case. If $\tau_2 > \tau_0$, it will be of aid to write (10) in the following form:

$$(10') \qquad \Delta S = \int_{\tau_1}^{\tau_0} (\delta - \delta_0)^2 d\tau + \int_{\tau_0}^{\tau_2} \delta^2 \, d\tau - 2\int_{\tau_0}^{\tau_2} \delta\delta_0 \, d\tau,$$

which shows that ΔS is again greater than zero, for the last integral on the right is negative since $\delta > 0$ and $\delta_0 \leqq 0$ from τ_0 to τ_2. The reasoning fails in this second part if the comparison curve is allowed to have a cusp through the change of sign of δ, and in what follows it will be shown how to find curves of this kind giving a smaller area than δ_0.

The Minimum Area as a Function of the End Inclinations. Suppose now that two minimizing curves, δ_1, δ_2, passing through the given end points have the same inclination τ_1 at the origin and the inclinations τ_2' and τ_2'', respectively, at P_2, and let the respective minimum areas be S_1 and S_2. Then $S_1 > S_2$ if $\tau_2' < \tau_2''$. This follows at once from the formula

$$\Delta S = \int_{\tau_1}^{\tau_2''} \delta_2^2\, d\tau - \int_{\tau_1}^{\tau_2'} \delta_1^2\, d\tau = -\int_{\tau_1}^{\tau_2'} (\delta_2 - \delta_1)^2 d\tau - \int_{\tau_2'}^{\tau_2''} \delta_2^2\, d\tau$$

which is obtained in identically the same way as (10), and, since no use is made of the fact that δ_1 is a minimizing curve, i.e., that

$$\delta_1 = (A_1 \tan \tau + B_1)/2,$$

it shows that the minimizing curve δ_2 gives a smaller area than *any other* curve through the same end points a..d having the same initial inclination but a final inclination less than or equal to that of the minimizing curve. A slightly different proof will also be given as it leads to an additional interesting result. Making use of the facts that $\delta_1 = (A_1 \tan \tau + B_1)/2$ and $\delta_2 = (A_2 \tan \tau + B_2)/2$, two equations similar to (9) may be written

$$\begin{aligned} S_1 &= \int_{\tau_1}^{\tau_2'} \delta_1^2\, d\tau = \int_{\tau_1}^{\tau_2''} \delta_1\delta_2\, d\tau, \\ S_2 &= \int_{\tau_1}^{\tau_2''} \delta_2^2\, d\tau = \int_{\tau_1}^{\tau_2'} \delta_1\delta_2\, d\tau. \end{aligned} \tag{11}$$

Hence

$$\Delta S = S_2 - S_1 = -\int_{\tau_2'}^{\tau_2''} \delta_1\delta_2\, d\tau. \tag{12}$$

Referring to the determination of A and B in (7) it will be seen that these two functions of τ_2 are continuous in τ_2 as long as $\tau_2 \neq \tau_1$. Hence δ is a continuous function of τ and τ_2, and it follows that, by taking $\tau_2'' - \tau_2'$ small enough, the sign of δ_2 can be made the same as that of δ_1 for all values of τ in the interval of integration, if $\delta_1 \neq 0$ for $\tau = \tau_2'$. It follows then that ΔS is negative. Since S is a continuous function of τ_2, which appears in the integrand as well as in the upper limit, it follows that S decreases even at points for which $\delta = 0$. The equation (12) leads by a simple reasoning to the result

$$\frac{dS}{d\tau_2} = -\delta^2 \tag{12'}$$

which also shows that S decreases as τ_2 increases. A similar analysis may be applied to the other extremity with the result that the area S decreases as τ_1 decreases.

The Symmetric Solution. If the two end points are taken on the same level, say $y_2 = 0$, and the inclinations at the ends are taken as the negatives of each other, $\tau_2 = -\tau_1 > 0$, then (7) shows that $A = 0$ and the equation (8) reduces to

$$(13)\qquad y = B \log \frac{\sec\left(\dfrac{x}{B} + \tau_1\right)}{\sec \tau_1} \qquad \begin{aligned} x_2 &= B(\tau_2 - \tau_1), \\ \delta &= B/2. \end{aligned}$$

In this case the curve is without a cusp and δ is a constant; and thus its caustic has a property somewhat similar to the tractrix. This curve is called the catenary of uniform strength. This is a solution of the problem in which the end conditions may be stated as follows. Given two vertical straight lines at $x = 0$ and $x = x_2$ and curves crossing these lines with the inclinations τ_1, τ_2, respectively, then (13) is the curve which gives the minimum area enclosed between it, its caustic, and the reflected rays at the crossing points. From the nature of the problem it will be seen that it is no real restriction to assume that all the curves pass through the origin. In this case the equation for y_2 in (3) drops out and the equation (4) becomes

$$S = \int_{\tau_1}^{\tau_2} \left(\delta - \frac{B}{2}\right)^2 d\tau - \int_{\tau_1}^{\tau_2} \left(\frac{B}{2}\right)^2 d\tau + \frac{B}{2} x_2,$$

and the minimum area is given by $\delta = B/2$ and has the value $Bx_2/4$, where B is to be determined from the equation for x_2 in (3).

This result may also be obtained by determining the value of y_2 which makes the minimum area in (6) attain its least value. Solving the equations (7) or (8) for A and B, we have

$$(14)\quad 4S = Ay_2 + Bx_2 = \frac{\left[y_2(\tau_2 - \tau_1) - x_2 \log \dfrac{\sec \tau_2}{\sec \tau_1}\right]^2}{-(\tau_2 - \tau_1)D_2} + \frac{x_2^2}{(\tau_2 - \tau_1)}.$$

Remembering that D_2 is negative it is clear that if

$$(15)\qquad y_2(\tau_2 - \tau_1) - x_2 \log \frac{\sec \tau_2}{\sec \tau_1} = 0,$$

S reaches its minimum value $x_2^2/4(\tau_2 - \tau_1)$. But the expression to the left in (15) is the value of $-D_2A$ and hence $A = 0$ gives the minimum area S.

THE POISSON INTEGRAL AND AN ANALYTIC FUNCTION ON ITS CIRCLE OF CONVERGENCE.

By A. Arwin.

Let $f(z)$ be an analytic function within the unit circle, having on the circumference C of this circle a finite number of singularities of logarithmic order, or of an order lower than that of a simple pole. Around the singular points on C we describe, in the interior of the circle of convergence, arcs of small circles of radii ϵ_p, and apply to these the process $\lim \epsilon_p \to 0$. We are then led to the conclusion that the integration of the Cauchy integral

$$f(\alpha) = \frac{1}{2\pi i}\int_c \frac{f(z)}{z-\alpha}\,dz \tag{1}$$

may be carried out over the singularities.

Let us consider the analytic function $f(1/\bar{z})/\bar{z}(\bar{z}-\bar{\alpha})$ for values $|\bar{z}| > 1$, $\bar{\alpha}$ being a point within the unit circle.

From Cauchy's theorem we have

$$\frac{1}{2\pi i}\int_c \frac{f(1/\bar{z})}{\bar{z}(\bar{z}-\bar{\alpha})}\,d\bar{z} = 0, \tag{2}$$

or

$$\frac{1}{2\pi i}\int_c \frac{f(z)}{\bar{z}-\bar{\alpha}}\,d\bar{z} - \frac{1}{2\pi i}\int_c \frac{f(z)}{\bar{z}}\,d\bar{z} = 0. \tag{2'}$$

From (1) and (2′) we get by subtraction

$$f(\alpha) = \frac{1}{2\pi}\int_c f(z)\left\{\frac{e^{i\theta}}{e^{i\theta} - Re^{i\psi}} + \frac{e^{-i\theta}}{e^{-i\theta} - Re^{-i\psi}}\right\}d\theta - \frac{1}{2\pi i}\int_c \frac{f(z)}{z}\,dz.$$

Placing $f(\alpha) = U(R, \psi) + iV(R, \psi)$, we have

$$\begin{aligned} U(R, \psi) &= \frac{1}{2\pi}\int_0^{2\pi} \frac{U(1, \theta)\{2 - 2R\cos(\psi-\theta)\}}{1 + R^2 - 2R\cos(\psi-\theta)}\,d\theta - \frac{1}{2\pi}\int_0^{2\pi} U(1, \theta)\,d\theta. \\ &= \frac{1}{2\pi}\int_0^{2\pi} \frac{U(1, \theta)(1 - R^2)}{1 + R^2 - 2R\cos(\psi-\theta)}\,d\theta, \end{aligned} \tag{3}$$

and a similar expression for $V(R, \psi)$. This is the well-known form of the integral of Poisson, except that now $U(1, \theta)$ may have logarithmic singularities, as well as algebraic singularities of an order lower than the first. When a singularity θ_1 is reached, we include this in an interval $\theta_1 - \epsilon$ to $\theta_1 + \epsilon$ and perform the operation $\lim \epsilon \to 0$. The integral over this in-

terval will then vanish. From formulas (1) and (2) we obtain the expression

$$\frac{1}{n!}f^{(n)}(0) = \frac{1}{2\pi}\int_0^{2\pi} f(z)\{e^{-in\theta} \pm e^{in\theta}\}d\theta,$$

where $f^{(n)}(0)$ denotes the nth derivative of $f(z)$ in the point $z = 0$.

Placing $f^{(n)}(0)/n! = \alpha_n + i\beta_n$, we have

$$\alpha_n = \frac{1}{\pi}\int_0^{2\pi} U(1, \theta) \cos n\theta d\theta = \frac{1}{\pi}\int_0^{2\pi} V(1, \theta) \sin n\theta d\theta,$$

$$\beta_n = \frac{1}{\pi}\int_0^{2\pi} V(1, \theta) \cos n\theta d\theta = -\frac{1}{\pi}\int_0^{2\pi} U(1, \theta) \sin n\theta d\theta,$$

which are the well-known values of the coefficients of the Fourier series.

If a point α be now moved into a regular point of $f(z)$ on the circumference of the circle of convergence, we shall have for this point

$$\text{(1')} \qquad \frac{1}{2}f(\alpha) = \frac{1}{2\pi i}\int_c \frac{f(z)}{z - \alpha}dz,$$

and

$$\text{(2'')} \qquad \frac{1}{2}f\left(\frac{1}{\bar{\alpha}}\right) = \frac{1}{2\pi i}\int_c \frac{f(z)}{\bar{z} - \bar{\alpha}}d\bar{z} - \frac{1}{2\pi i}\int_c \frac{f(z)}{\bar{z}}d\bar{z},$$

from which follows

$$\begin{aligned} 0 = \frac{1}{2\pi}\int_0^{2\pi} f(z)\left\{\frac{e^{i\theta}}{e^{i\theta} - e^{i\psi}} + \frac{e^{-i\theta}}{e^{-i\theta} - e^{-i\psi}}\right\}d\theta \\ - \frac{1}{2\pi}\int_0^{2\pi}\{U(1, \theta) + iV(1, \theta)\}d\theta \\ = \frac{1}{2\pi}\int_0^{2\pi} f(z)\left[1 + \sum_1^n e^{im(\psi-\theta)} + 1 + \sum_1^n e^{-im(\psi-\theta)}\right]d\theta + \\ + \frac{1}{2\pi}\int_0^{2\pi} f(z)\left\{\frac{e^{i(n+1)(\psi-\theta)}}{1 - e^{i(\psi-\theta)}} + \frac{e^{-i(n+1)(\psi-\theta)}}{1 - e^{-i(\psi-\theta)}}\right\}d\theta \\ - \frac{1}{2\pi}\int_0^{2\pi}\{U(1, \theta) + iV(1, \theta)\}d\theta, \end{aligned}$$

or

$$\begin{aligned} 0 = \frac{1}{2\pi}\int_0^{2\pi} f(z)d\theta + \sum_1^n \frac{1}{\pi}\int_0^{2\pi} f(z) \cos m(\psi - \theta)d\theta \\ + \frac{i}{2\pi}\int_0^{2\pi} f(z)\left\{\frac{e^{i(n+1)(\psi-\theta)}}{2 \sin\frac{\psi - \theta}{2}e^{i\frac{\psi-\theta}{2}}} - \frac{e^{-i(n+1)(\psi-\theta)}}{2 \sin\frac{\psi - \theta}{2}e^{-i\frac{\psi-\theta}{2}}}\right\}d\theta. \end{aligned}$$

That is

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{U(1,\,\theta)\sin(2n+1)\dfrac{\psi-\theta}{2}}{\sin\dfrac{\psi-\theta}{2}}d\theta$$

$$=\frac{1}{2\pi}\int_0^{2\pi}U(1,\,\theta)d\theta+\sum_1^n\frac{1}{n}\int_0^{2\pi}U(1,\,\theta)\cos m(\psi-\theta)d\theta. \quad (4)$$

This is the familiar summation formula for the common Fourier series. A similar expression is obtained for $V(1, \theta)$.

Since for a value θ_1 of θ $U(1, \theta)$ can have only a singularity of lower order than the extension of the interval of integration ϵ, or only a singularity of an order lower than the linear, we may apply the general theory of Fourier series to the integral on the left hand side of equation (4). We have then for every regular place ψ of $U(1, \theta)$

$$(5) \qquad U(1,\,\psi)=\lim_{n\to\infty}\frac{1}{2\pi}\int_0^{2\pi}\frac{U(1,\,\theta)\sin(2n+1)\dfrac{\psi-\theta}{2}}{\sin\dfrac{\psi-\theta}{2}}d\theta.$$

An addition of (1′) and (2″) would, on account of (4) and (5), have led to the formula

$$(6) \qquad 0=\lim_{n\to\infty}\int_0^{2\pi}\frac{U(1,\,\theta)\cos(2n+1)\dfrac{\psi-\theta}{2}}{\sin\dfrac{\psi-\theta}{2}}d\theta$$

which could also have been proved directly.

From these results we conclude conversely that an analytic function which has only a finite number of singularities on its circle of convergence, these singularities being of logarithmic order or of order lower than that of a simple pole, may be represented in every regular point by the familiar series which is derived by means of Cauchy's integral and which is valid within the circle of convergence. This fact, it seems, is equivalent to the contents of a theorem by Fatou-M. Riesz.*

LUND, SWEDEN,
June, 1920.

* E. Landau, Darstellung und Begründung einiger neuerer Ergebnisse der Funktionentheorie. Berlin, 1916.

SYSTEMS OF CIRCUITS ON TWO-DIMENSIONAL MANIFOLDS.

By H. R. Brahana.

1. In this paper we first give a method of reducing any two-dimensional manifold to one of the known polygonal normal forms. The method used is one by which a polygon on which the manifold is represented is subjected to a series of transformations by cutting it apart in a simple manner and then joining it together again so as to obtain a new polygon representing the same manifold.

We next (§§ 11 to 18) apply the same series of transformations to the problem of reducing a system of curves on the manifold to a normal form.* We then introduce certain matrices of separation by means of which the relations among the pairs of sides of the polygon are described and study the effect on these matrices of the transformation of cutting. By this means we obtain a number of theorems on systems of curves which follow closely along the lines of the theory indicated in Poincaré's "Cinquième Complement à l'Analysis Situs."†

We shall use the terms *manifold, cell, circuit, orientable, one-sided*, etc., as they are defined by Professor Veblen in his Cambridge Colloquium lectures on Analysis Situs. It is there shown (Chapt. II, § 65) that any two-dimensional manifold can be imaged on a planar polygon in such a way that any point of the manifold has for its image an interior point, a pair of "conjugate points" (cf. § 3 below), or a "conjugate set of vertices" of the polygon.

I take this opportunity to acknowledge my indebtedness to Dr. J. W. Alexander for suggestions and to Professor O. Veblen for proposing the problem and for advice in working it out.

2. **Conjugate Points and Sides of a Polygon.** Consider a polygon of an even number, $2n$, of sides in a Euclidean plane. Let P_1, P_2, P_3 be three distinct points taken in the order $P_1P_2P_3$ on the side a_i of the polygon. These three points determine a sense of description of the boundary of the polygon. A (1-1) continuous correspondence may be set up between the points of a_i and the points of any other side a_j of the polygon. Let such a correspondence be established and let the points which correspond to $P_1P_2P_3$ be $P_1'P_2'P_3'$ respectively. In case the three points $P_1'P_2'P_3'$ determine the same sense on the boundary of the polygon as is determined

* This question was first considered by Jordan, Journal de math., (2) 11, pp. 105, 110.

† Rendiconti del Circolo Matematico di Palermo, vol. 18 (1904), p. 45.

by the points $P_1P_2P_3$, the correspondence will be called *direct;* in case the two senses are not the same, the correspondence will be called *opposite.*

Suppose the sides of the polygon have been paired arbitrarily and denote the members of a pair by a_i and a_i'. Let a_i be called the side *conjugate* to the side a_i', and a_i' the side conjugate to a_i. Let a correspondence, direct or opposite, be established between the members of each pair. Two corresponding points P_1 and P_1', interior to a_i and a_i' respectively, will be called a *conjugate pair* of points.

3. Choose $4n$ points on the boundary of the polygon in the following manner: Take two arbitrary distinct points on each of the n sides a_i; then take the two points conjugate to them on each of the n sides a_i'. (Fig. 1.) Let the two points nearest to the vertex P_i, one on each of the

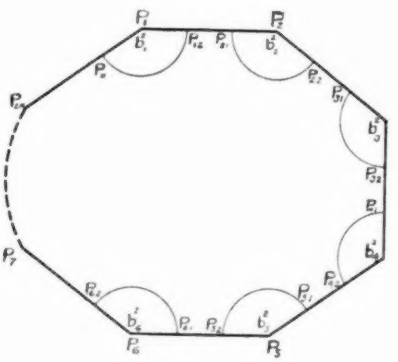


FIG. 1.

sides that has an end at P_i, be called P_{i1} and P_{i2}. Join P_{i1} to P_{i2} by a 1-cell p_i on the polygon. Do the same for each vertex, choosing the 1-cells p_i so that no two intersect. Let the 2-cell whose boundary is made up of the segments $P_{i1}\ P_i$ and $P_{i2}\ P_i$, the 1-cell p_i, and the points P_{i1}, P_{i2}, and P_i be called b_i^2. Consider the side $P_{i1}\ P_i$ of the 2-cell b_i^2. There is a unique 2-cell b_j^2 one of whose sides $P_{j1}\ P_j$ (or $P_{j2}\ P_j$) is a segment conjugate to the segment $P_{i1}\ P_i$. Join together these 2-cells by matching up conjugate points on their boundaries. Then there exists a unique 2-cell b_k^2 one of whose sides is conjugate to $P_{j2}\ P_j$ (or $P_{j1}\ P_j$). Join b_k^2 to b_j^2 in the same manner. This may be continued until a 2-cell b_l^2 is reached one of whose sides is the conjugate of the side $P_{i2}\ P_i$ of the 2-cell b_i^2. The vertices $P_i\ P_j\ P_k\ \cdots\ P_l$ of the polygon which are on the boundaries of such a set of 2-cells will be called a *conjugate set of vertices.*

4. If the 2-cells b_i^2, b_j^2, $\cdots$, b_l^2 which determine a conjugate set of vertices be fitted together at their edges in such a way that conjugate pairs of points coincide, it is evident that they will constitute a single

2-cell. Hence it is evident that, for any polygon of $2n$ sides on which conjugate pairs of points and sets of vertices have been defined, there can be found a two-dimensional manifold such that there is a continuous correspondence in which each point of the polygon corresponds to one, and only one, point of the manifold and each point of the manifold corresponds either to one, and only one, point interior to the polygon, or to a pair of conjugate points on the boundary, or to a set of conjugate vertices. Conversely, for any two-dimensional manifold a polygon of $2n$ sides can be found (cf. the reference above) which is its image in the manner just described.

5. We shall assume that a sense has been arbitrarily assigned to each of the sides a_i. This sense may be denoted by the order of any three distinct points on a_i. The three conjugate points on a_i' determine a definite sense on a_i'. In case the senses of a_i and a_i' for all values of i are such that one of them agrees and the other disagrees with a fixed sense of description of the boundary of the polygon, it is obvious that the manifold represented by the polygon is *orientable* or *two-sided*. In case there is one pair of sides a_i and a_i' the senses of which both agree with a fixed sense of description of the boundary of the polygon, it is equally obvious that the manifold represented is *one-sided*.

6. **Transformations of the Polygon.** A 1-cell x on the polygon with its ends on the boundary divides the polygon into two 2-cells α and β (see Fig. 3). Suppose the side b_2 is on the boundary of α and the side b_2' is on the boundary of β. By cutting the polygon along x and joining the two 2-cells by matching up conjugate points of the two sides b_2 and b_2' a new polygon is obtained (see Fig. 4) which is in the same relation to the manifold as was the original polygon. If c is the image on the manifold of the 1-cell x, then on the new polygon the image of c will be two conjugate sides; the image of a point interior to c will be a pair of conjugate points.

This transformation will be referred to as the *method of cutting*. The 1-cell x will be called a *cut*. The method of cutting will now be used to reduce the polygon to a normal form.* We shall first reduce to one the number of points a_i^0 of the manifold which correspond to vertices of the polygon, and secondly shall obtain a definite arrangement of pairs of conjugate sides of the polygon.

7. **Reduction to a Single Conjugate Set of Vertices.** A sense may be assigned arbitrarily to each of the edges a_i and denoted by the order of any three distinct points on it. The three conjugate points on a_i' deter-

* The application of the method of cutting to the normalization of a polygon is due to Professor Veblen; it was first given by him in a seminar on Analysis Situs in 1915.

mine a sense on a_i'. The sense of any side determines a sense of description of the boundary of the polygon.

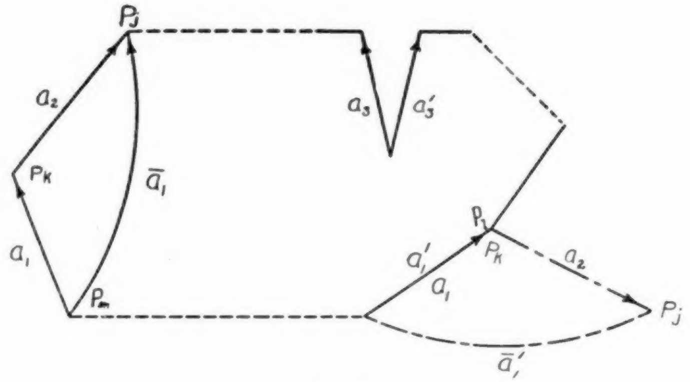


Fig. 2.

Reduction 1. If vertices of the polygon correspond to more than one point of the manifold, there will be some side, say a_2, whose ends, P_j and P_k, correspond to distinct points of the manifold; let a_1 be a side with one end at P_k and the other at a vertex P_m (Fig. 2). First let us suppose that the side a_2 is not a_1'. Let P_l be the end of a_1' which corresponds to the same point of the manifold as P_k. Draw a cut $\bar{a}_1$ joining P_m to P_j and join the two parts of the polygon along the sides a_1 and a_1'. This gives a polygon on which the number of vertices in the conjugate set to which P_k and P_l belong has been reduced by one; the number of sides of the polygon has not been changed.

Reduction 2. In case a side, say a_3 (Fig. 2), joins two vertices which correspond to different poïnts of the manifold and has an end in common with its conjugate side a_3', we have the case excluded in Reduction 1. From the way in which points of a_3 and a_3' correspond it follows that a_3 and a_3' must be oppositely sensed. Hence by coalescing the pairs of conjugate points of a_3 and a_3' a polygon can be formed from which the two sides a_3 and a_3' and their common vertex have been removed. The number of points of the manifold to which vertices of the polygon correspond has been reduced by one.

8. These reductions may be continued so long as there is more than one point of the manifold to which vertices of the polygon correspond. By each step either a conjugate set of vertices is removed, or the number of vertices in one conjugate set is increased while the number of vertices in another conjugate set is reduced by one (Reduction 1); also the conjugate set of which the number of vertices is to be increased can be chosen arbitrarily, because the rôles of P_j and P_k may be interchanged in Reduc-

tion 1. Hence by a finite number of steps a polygon may be obtained whose vertices constitute a single conjugate set corresponding to an arbitrarily chosen 0-cell a_i^0 of the manifold, or else a polygon of two sides may be obtained whose vertices constitute two conjugate sets, and whose sides are oppositely sensed. The manifold defined by the latter polygon is a sphere.

Hereafter we shall call the 0-cell a_i^0 the point A. Each pair of conjugate sides of the polygon will be imaged on a 1-cell on the manifold whose ends coincide with A. In other words, each pair of conjugate sides of the polygon will correspond to a simple circuit on the manifold through the point A.*

9. **Normalization of the Two-Sided Polygon.** Let us first consider the two-sided case and show how to obtain a group $x\ y\ x'\ y'$ of four consecutive sides on the boundary of the polygon. Draw a cut x joining the two forward ends of a_i and a_i' (a_2 and a_2' in Fig. 3). Let the two parts of the

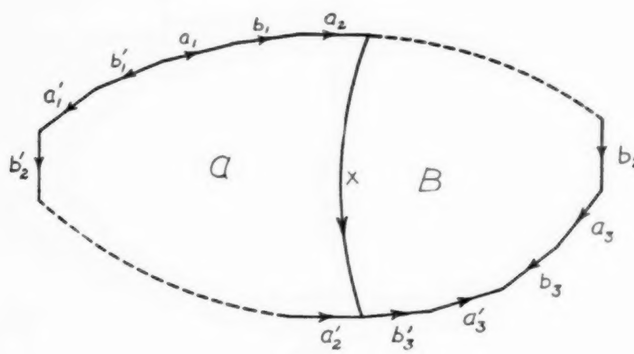


FIG. 3

polygon be α and β where a_i and a_i' are on α. There must be some side a_j (b_2' in Fig. 3) on α whose conjugate a_j' is on β, otherwise the vertices of β together with the two vertices of α at the forward ends of a_i and a_i' would constitute a conjugate set without including all the vertices of the polygon. Join α and β along the sides a_j and a_j'. On the resulting polygon the three sides $a_i\ x\ a_i'$ will be consecutive (Fig. 4). Draw a cut y joining the forward ends of x and x'. Join the two parts of the polygon along the sides a_i and a_i' (Fig. 5). The four sides $y'\ x\ y\ x'$ are consecutive and in that order.

This process may be repeated for any other pair of conjugate sides a_k and a_k' without disturbing the arrangement of the sides $x\ y\ x'\ y'$ for no cut will be drawn from a vertex at which two of these sides abut.

* The same result could be obtained by shrinking to points 1-cells joining distinct 0-cells of the manifold.

From the above reasoning it follows that the number of sides of the polygon of a two-sided manifold, if the polygon has a single conjugate set of

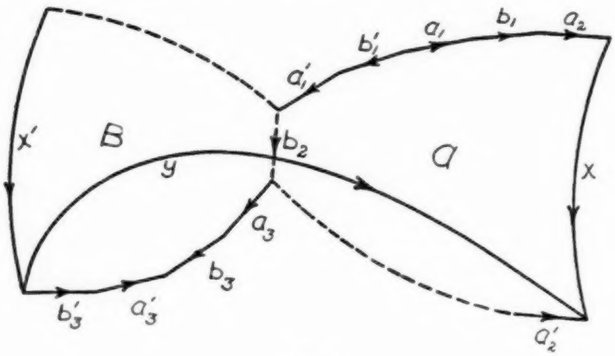


FIG. 4

vertices, is a multiple of four. Completing the reduction and changing the notation we get the following arrangement of the sides of the polygon:

$$a_1\ b_1\ a_1'\ b_1'\ a_2\ b_2\ a_2'\ b_2'\ \cdots\ a_p\ b_p\ a_p'\ b_p'.$$

This is the normal form of the polygon. The number p is called the *genus* of the manifold. The *connectivity* R_1 of the manifold is $2p + 1$.

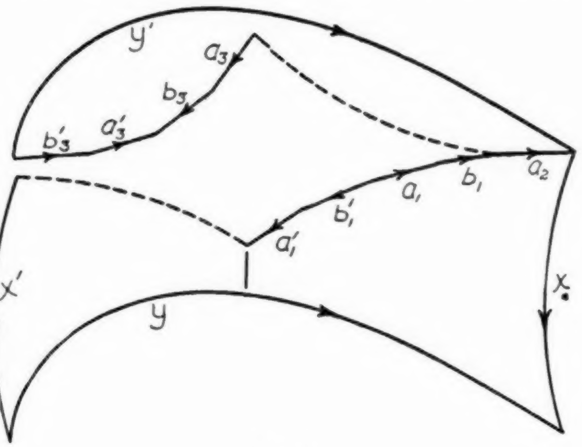


FIG. 5.

10. **Normalization of the One-Sided Polygon.** In the consideration of the one-sided case we make use of the transformation just described for the two-sided case if there exists a group of four sides having the same relations among themselves that the sides a_i, a_i', a_j, and a_j' had above. Thus we obtain on the boundary of the polygon a certain number of groups of four consecutive sides in the order $a_i\ b_i\ a_i'\ b_i'$.

Let a_k and a_k' be a conjugate pair of sides which have the same sense (a_3 and a_3' in Fig. 6). Draw a cut x joining the forward ends of a_k and

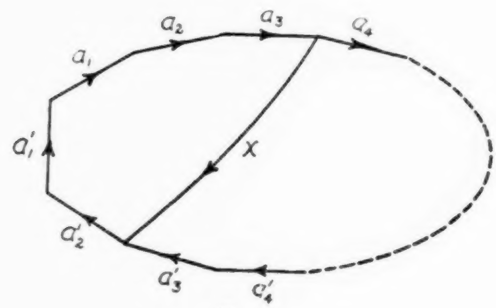


FIG. 6.

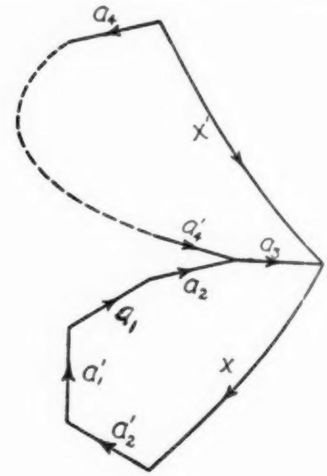


FIG. 7.

a_k', and join the parts of the polygon along a_k and a_k'. This replaces the pair $a_k\ a_k'$ by the pair $x\ x'$ (Fig. 7) which is a pair of consecutive conjugate sides having the same sense. By application of the two transformations the sides of the polygon may be arranged in groups of four of the form $a_i\ b_i\ a_i'\ b_i'$ and groups of two of the form $c_j\ c_j'$.*

A group of six sides of the form $a_i\ b_i\ a_i'\ b_i'\ c_j\ c_j'$ may be replaced by three groups of two of the form $c_k\ c_k'\ c_l\ c_l'\ c_m\ c_m'$. Draw a cut x joining the forward ends of a_i and c_j (Fig. 8). Join the two parts of the polygon

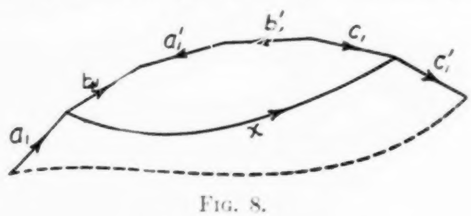


FIG. 8.

along the sides c_j and c_j'. This gives six consecutive sides $a_i\ x\ b_i'\ a_i'\ b_i\ x'$ (Fig. 9). Draw a cut y joining the backward end of a_i to the forward

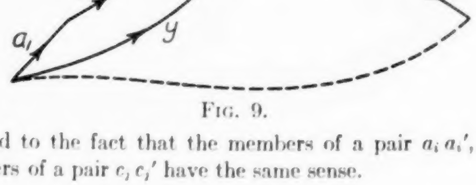


FIG. 9.

* Attention is called to the fact that the members of a pair $a_i\,a_i'$, or $b_i\,b_i'$, are oppositely sensed, and that members of a pair $c_j\,c_j'$ have the same sense.

end of b_i', and join the two parts of the polygon along the sides a_i and a_i'. This gives the six consecutive sides $y\ y'\ b_i'\ x\ b_i\ x'$ (Fig. 10). Draw a cut

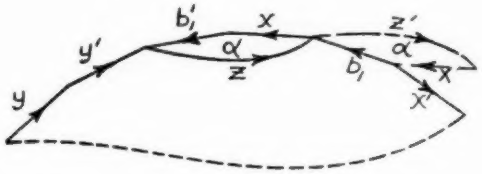


FIG. 10.

z joining the forward ends of b_i and b_i', join the two parts of the polygon along the sides b_i and b_i'. This gives the six consecutive sides $y\ y'\ z\ z'\ x\ x'$, which is the desired form (Fig. 10).

From the above it follows that the polygon of a one-sided manifold may be put in the form:

$$(1) \qquad a_1\, a_1'\, a_2\, a_2'\, a_3\, a_3' \cdots a_{R_1-1}\, a_{R_1-1}'.$$

The number R_1 is the *connectivity* of the manifold.

By applying the inverse of the reduction just described to a set of three consecutive pairs the polygon of a one-sided manifold may be put in one of the two forms:

$$(a) \qquad a_1\, b_1\, a_1'\, b_1'\, a_2\, b_2\, a_2'\, b_2' \cdots a_p\, b_p\, a_p'\, b_p'\, c_1\, c_1';$$

or

$$(b) \qquad a_1\, b_1\, a_1'\, b_1'\, a_2\, b_2\, a_2'\, b_2' \cdots a_p\, b_p\, a_p'\, b_p'\, c_1\, c_1'\, c_2\, c_2',$$

according as $R_1 - 1$ is odd or even.

11. **Fundamental Sets of Circuits.**[1] When the polygon has been so transformed that the vertices constitute a single conjugate set the image on the manifold of a pair of conjugate sides of the polygon is a simple circuit through the point A. No two of these circuits have any other point in common. The circuits constitute the complete boundary of a 2-cell which contains all the points of the manifold which are not on the circuits. Such a set of circuits has been called by Poincaré a *fundamental set*.

The discussion in the first part of this paper proves the existence of a fundamental set. We shall now prove that a fundamental set can be obtained with an arbitrary point A_1 of the manifold as the point A. If the image P_1 of A_1 is interior to the polygon, draw an arc p connecting P_1 with some vertex P of the polygon. Cut the polygon along the arc p. This gives a polygon with two more sides than the original polygon and with two conjugate sets of vertices. Now apply Reduction 1 of § 7 in such a way that the number of vertices in the conjugate set which corre-

sponds to A_1 is increased. This may be continued until by application of Reduction 2 (§ 7) the conjugate set which corresponds to A is removed, and the number of sides of the polygon is reduced by two. This gives a polygon of the same number of sides as the original one and with a single conjugate set of vertices. Consequently we have a new fundamental set of circuits, each passing through A_1, and the number of circuits in this set is the same as in the original set.

If the point P_1 were on a side of the polygon, the number of sides would be increased by two if we considered P_1 and its conjugate point P_1' as vertices. The above procedure could then be carried out giving the same result.

12. In considering a simple circuit $\bar{C}$ on the manifold we may assume, as a result of what has just been proved, that the point A of a fundamental set F is on the circuit. Let us consider the polygon whose conjugate pairs of sides are imaged on the circuits of F, and let us suppose that $\bar{C}$ has a finite number of points in common with circuits of F.* The image of $\bar{C}$ on the polygon will be a set of arcs $[C_i']$. If $\bar{C}$ has no point in common with F other than A, this set will consist of a single arc having its ends at two vertices of the polygon; these two vertices will be distinct unless $\bar{C}$ divides the manifold into two parts. If $\bar{C}$ has points other than A in common with F, two of the arcs $[C_i']$ will have one end each at a vertex of the polygon; the other ends of arcs of $[C_i']$ will be at points interior to the sides of the polygon. The second case may be reduced to the first by a proper choice of the fundamental set F; this may be done by the method of cutting.

For, let C_1' be an arc with one end at the vertex P_i and the other at a point P interior to the side a_i (Fig. 11). Draw a cut x joining P_i to

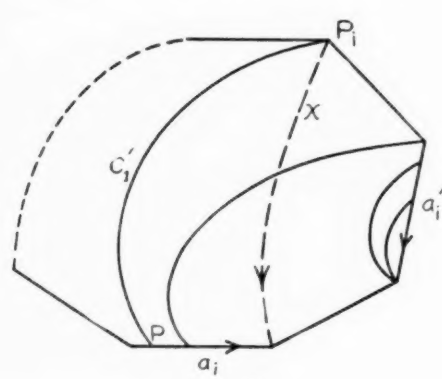


Fig. 11.

* A fundamental set may always be chosen so that the above condition is satisfied.

an end of a_i such that a_i and a_i' are on different parts of the polygon. The cut x can be drawn so that it has no intersections with C_1', so that it has no intersections with any arc joining two boundary points neither of which is interior to a_i, and so that it has no more than one intersection with any arc having an end on a_i. By joining the two parts of the polygon along the sides a_i and a_i' a new polygon is obtained (Fig. 12) such that

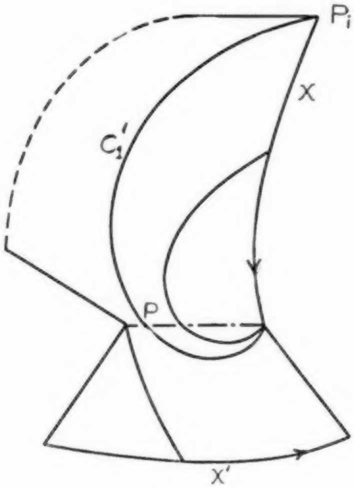


FIG. 12.

the number of ends of arcs at points interior to the sides is at least one less than on the original polygon. This process may be continued until this number is zero, i.e., until a polygon is obtained on which the image of C is a single arc C_i joining two vertices.

13. Let us suppose that the circuit $\bar{C}$ is not homologous to zero, i.e., that it does not divide the manifold into two parts. Then the two ends of the arc C_i are distinct; and if α and β are the two parts of the polygon determined by C_i, there must be some side a_i on the boundary of α whose conjugate side a_i' is on the boundary of β. Hence, if we cut the polygon along C_i and join the two parts along a_i and a_i', a polygon is obtained on which the image of $\bar{C}$ is an arc joining two consecutive vertices. Hence, *any simple circuit which is not homologous to zero may be made a member of a fundamental set.*

14. **Relations between Two Fundamental Sets.** To compare two fundamental sets F and F_1 we may assume that the points A and A_1 coincide. Let conjugate pairs of sides of the polygon be images of circuits of F. No circuit or set of circuits of F_1 divides the manifold into two parts. The image of F_1 on the polygon will be a set of non-intersecting arcs

$[C_i']$ having their ends on the boundary. By the method of § 12 we may obtain a polygon on which one of the circuits of F_1 is imaged on an arc C_i joining two vertices. Also since no two of the arcs $[C_i']$ intersect, we may obtain by the same method a polygon on which a second circuit of F_1 is imaged on an arc C_j joining two vertices; for since neither of the ends of C_i is interior to any side of the polygon, none of the required cuts will cross C_i. Continuing this process a polygon is obtained on which the image of F_1 is a set of arcs $[C_i]$ each having its ends at vertices of the polygon.

We will next see how a polygon may be obtained which is such that each conjugate pair of sides corresponds to a circuit of F_1, and which is such that every circuit of F_1 corresponds to a pair of conjugate sides. It will also be seen that the number of sides of this polygon is the same as the number of sides of the original polygon.

If C_i is an arc joining the two ends of a_i, cut the polygon along C_i and join the two parts along the sides a_i and a_i'. This gives a conjugate pair of sides whose image on the manifold is a circuit of F_1. There exists no arc C_j joining the ends of a side of this conjugate pair, for if there were such an arc, it and C_i would divide the manifold into two regions.

Let the transformation described in the last paragraph be carried out for each of the arcs of $[C_i]$ which joins two consecutive vertices of the polygon. If C_j is an arc which joins two vertices of the polygon which are not consecutive, it divides the polygon into two parts, α and β, and there must exist a conjugate pair of sides a_i and a_i' of which one is on the boundary of α and the other is on the boundary of β, and which is not the image of any of the circuits of F_1; otherwise any arc on the manifold joining two points $\overline{P}_\alpha$ and $\overline{P}_\beta$ would intersect one of the circuits of F_1. Cutting the polygon along the arc C_j and joining the two parts along the sides a_i and a_i', a polygon is obtained which has a conjugate pair of sides whose image on the manifold is a circuit of F_1. By the above methods a polygon may be obtained which has a pair of conjugate sides for every circuit of F_1. It remains to be seen that every pair of conjugate sides of this polygon is imaged on a circuit of F_1. If this were not so, F_1 would not bound a 2-cell and so would not be a fundamental set.

15. **Invariance of the Connectivity.** Since none of the transformations used changes the number of sides of the polygon and since the normalization of a polygon whose vertices constitute a single conjugate set does not change the number of sides, it follows that the values of the connectivity determined by the two fundamental sets are the same. The connectivity is independent of the particular fundamental set in terms of which it was defined.

16. **Equivalences and Homologies.** The transformations involved in

what we have called the method of cutting amount in every case to replacing one of a set of circuits by a new circuit which is related to the circuits of the original set by an equivalence in the sense of Poincaré.* For example, in Fig. 8 the cut x joins the rear end of b_1 to the front end of c_1 and we have

$$x^1 \equiv b_1^1 - a_1^1 - b_1^1 + c_1^1$$

because b_1^1, $-a_1^1$, $-b_1^1$, c_1^1 and $-x^1$ taken in order bound a 2-cell. The two parts of the polygon are joined together along c_1 and c_1' (see Fig. 9) so that the set of circuts a_1^1, b_1^1, c_1^1, $\cdots$ has been converted into $a_1^1\ b_1^1\ x_1^1$ $\cdots$ where the two sets of curves are related by the set of equivalences:

$$\begin{aligned}
a_1^1 &\equiv a_1^1 \\
b_1^1 &\equiv b_1^1 \\
x^1 &\equiv b_1^1 - a_1^1 - b_1^1 + c_1^1 \\
d_1^1 &\equiv d_1^1 \\
&\cdot \\
&\cdot \\
&\cdot
\end{aligned}$$

In case the vertices of the polygon are all in one conjugate set, these equivalences are what Poincaré calls *proper* equivalences because all the 1-cells in question begin and end at the same point. In case they are not all in one set, the equivalences are what he calls *improper* equivalences.

In the general case it is clear that, if we pass by the method of cutting from a polygon whose sides represent a set of 1-cells $a_1, a_2, \cdots, a_m$ to one whose sides represent a set of 1-cells $b_1, b_2, \cdots, b_m$, we have a set of equivalences of the form

$$(1)\qquad \begin{aligned}
b_1 &\equiv \epsilon_1^{11}a_1 + \epsilon_1^{12}a_2 + \cdots + \epsilon_1^{1m}a_m + \epsilon_1^{21}a_1 + \epsilon_1^{22}a_2 + \cdots + \epsilon_1^{km}a_m \\
b_2 &\equiv \epsilon_2^{11}a_1 + \epsilon_2^{12}a_2 + \cdots + \epsilon_2^{1m}a_m + \epsilon_2^{21}a_1 + \epsilon_2^{22}a_2 + \cdots + \epsilon_2^{km}a_m \\
&\cdot \qquad \cdot \qquad\qquad\qquad \cdot \qquad\qquad\qquad \cdot \\
&\cdot \qquad \cdot \qquad\qquad\qquad \cdot \qquad\qquad\qquad \cdot \\
&\cdot \qquad \cdot \qquad\qquad\qquad \cdot \qquad\qquad\qquad \cdot \\
b_n &\equiv \epsilon_n^{11}a_1 + \epsilon_n^{12}a_2 + \cdots + \epsilon_n^{1m}a_m + \epsilon_n^{21}a_1 + \epsilon_n^{22}a_2 + \cdots + \epsilon_n^{km}a_m
\end{aligned}$$

in which the ϵ's are $+1$, -1, or 0.

17. The terms of an equivalence are not commutative. If we treat them as if they were commutative and collect terms, the equivalences (1) reduce to the *homologies*

$$(2)\qquad \begin{aligned}
b_1 &\sim \eta_1^1a_1 + \eta_1^2a_2 + \eta_1^3a_3 + \cdots + \eta_1^m a_m \\
b_2 &\sim \eta_2^1a_1 + \eta_2^2a_2 + \eta_2^3a_3 + \cdots + \eta_2^m a_m \\
&\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\
b_n &\sim \eta_n^1a_1 + \eta_n^2a_2 + \eta_n^3a_3 + \cdots + \eta_n^m a_m,
\end{aligned}$$

* Loc. cit., p. 60; see also Veblen, loc. cit., Chap. V, § 28.

in which the η's are integers. It is easily seen that in a homology the right and left sides together constitute the boundary of an oriented two-dimensional manifold though not in general a 2-cell.

If the coefficients η of these homologies are reduced modulo 2, we obtain the following homologies:

$$
\begin{array}{l}
b_1 \sim \zeta_1{}^1 a_1 + \zeta_1{}^2 a_2 + \cdots + \zeta_1{}^m a_m \\
b_2 \sim \zeta_2{}^1 a_1 + \zeta_2{}^2 a_2 + \cdots + \zeta_2{}^m a_m \\
\cdot \qquad \cdot \qquad \cdot \\
\cdot \qquad \cdot \qquad \cdot \\
\cdot \qquad \cdot \qquad \cdot \\
b_n \sim \zeta_n{}^1 a_1 + \zeta_n{}^2 a_2 + \cdots + \zeta_n{}^m a_m
\end{array} \pmod 2 \tag{3}
$$

in which the ζ's are all 1 or 0. It is easily seen that in a homology (mod 2) the right and left sides constitute the boundary of a two-dimensional manifold which need not be oriented. (See Veblen, loc. cit., Chap. II, § 37.)

It is obvious that the homologies (mod 2) are the simplest and easiest to work with, that the Poincaré homologies are the next simplest, and that the equivalences are the most difficult on account of their non-commutative character. We shall therefore in what follows first consider the homologies (mod 2), then the Poincaré homologies.

18. We have now seen that it is possible to pass from any fundamental set of circuits to any other by the method of cutting, and also that the number of circuits in all fundamental sets is the same. In terms of the equivalences of § 16 this means that, between any two fundamental sets $a_1\, a_2 \cdots a_\mu$ and $\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_\mu$, there exist the equivalences

$$
\begin{array}{l}
\bar{a}_1 \equiv \sum_{i=1}^{m} \sum_{j=1}^{\mu} \epsilon_1{}^{ij} a_j \\
\cdot \qquad \cdot \\
\cdot \qquad \cdot \\
\cdot \qquad \cdot \\
\bar{a}_\mu \equiv \sum_{i=1}^{m} \sum_{j=1}^{\mu} \epsilon_\mu{}^{ij} a_j
\end{array}
$$

From this there follow the homologies

$$\bar{a}_p \sim \sum_{j=1}^{\mu} \beta_p{}^j a_j \qquad (p = 1, 2, \cdots, \mu),$$

where

$$\beta_p{}^j = \sum_{i=1}^{m} \epsilon_p{}^{ij}.$$

We now want to investigate the question as to what are the conditions under which two fundamental sets of circuits satisfy a set of equivalences.

$$\bar{a}_p \equiv a_p \qquad (p = 1, 2, \cdots, \mu).$$

This is related to the question as to whether they satisfy the much weaker conditions

$$\bar{a}_p \sim a_p,$$

or the still weaker condition

$$\bar{a}_p \sim a_p \pmod 2.$$

With a view to studying these questions we introduce certain matrices expressing the relations among the circuits of a fundamental set.

19. **The Separation Matrix.** Suppose that each side of the polygon has been given a sense in the manner described in § 5. A 1-cell joining the forward ends of a_i and a_i' divides the polygon into two parts α and β. If one and only one of the sides a_j and a_j' is on the boundary of α, we will say that the conjugate pair $a_j\, a_j'$ *separates* $a_i\, a_i'$. As an obvious consequence of the definition we get the following theorems:

1: *If the pair $a_j\, a_j'$ separates the pair $a_i\, a_i'$, then the pair $a_i\, a_i'$ separates the pair $a_j\, a_j'$*;

2: *If the two sensed sides a_i and a_i' determine the same sense of description of the boundary of the polygon, then the pair $a_i\, a_i'$ separates itself*; *in the opposite case the pair $a_i\, a_i'$ does not separate itself.*

20. We will now construct a square matrix of $R_1 - 1$ rows which is uniquely determined by the polygon. Let e_{ij}, the element in the ith row and the jth column, be 1 or 0 according as the pair $a_j\, a_j'$ separates or does not separate the pair $a_i\, a_i'$; this matrix will be called the *separation matrix* of the polygon.

From the first theorem of § 19 it follows that e_{ij} is equal to e_{ji}; and from the second it follows that e_{ii} is 1 or 0 according as the sides a_i and a_i' have the same or opposite senses.

21. The separation matrix of the normalized polygon of a two-sided manifold is the following:

$$\begin{Vmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{Vmatrix}$$

The separation matrix of the polygon of a one-sided manifold in the normal form (1) of § 10 is:

$$\begin{array}{cccccccc} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{array}$$

These two matrices are also the separation matrices of the polygons whose sides are respectively in the order:

$$a_1\, b_1\, a_2\, b_2\, a_3\, b_3\, a_3'\, b_3'\, a_2'\, b_2'\, a_1'\, b_1' \text{ (Fig. 3)};$$

and

$$a_1\, a_2\, a_3\, a_4\, a_4'\, a_3'\, a_2'\, a_1' \text{ (Fig. 6)}.$$

Thus we see that a given separation matrix corresponds in general to more than one polygon. We will return later to the relations between two polygons which have the same separation matrix.

22. Let us first consider the effect of cutting along a 1-cell $\bar{a}_1$ equivalent to $a_1 + a_2$ and joining the two parts together along the sides a_1 and a_1' (cf. Fig. 2). This amounts to changing the fundamental set by the equivalence transformation

$$\begin{aligned} \bar{a}_1 &\equiv a_1 + a_2 \\ \bar{a}_2 &\equiv a_2 \\ \cdot \quad & \quad \cdot \\ \cdot \quad & \quad \cdot \\ \cdot \quad & \quad \cdot \\ \bar{a}_\mu &\equiv a_\mu . \end{aligned}$$

We shall see that this changes the polygon π into a new polygon whose separation matrix is obtained from that of π by multiplying on the right by the matrix of the above transformation and on the left by the conjugate of that matrix, and then reducing each element modulo 2.

23. Let us consider first the case where a_1 and a_1' have opposite senses on the boundary of the polygon. On comparing the separation matrix of the new polygon with that of the old we see: (*a*) The first row and column are unchanged—i.e., the row and column corresponding to $\bar{a}_1\, \bar{a}_1'$ on the transformed matrix M_1 are the same as the row and column corresponding to $a_1\, a_1'$ on the original matrix M; (*b*) The second row and the second column of M_1 are the result of adding the first row of M to the second row, adding the first column to the second column, and reducing each element modulo 2. For if the element e_{2i} of M_1 is 1, a single side of the pair $a_i\, a_i'$ is on each part of the boundary of π between a_1' and a_2'.

Hence $a_i\,a_i'$ separates one but not both of the pairs $a_1\,a_1'$ and $a_2\,a_2'$, and hence just one of the elements e_{1i} and e_{2i} of M is 1. Conversely, if one and only one of the elements e_{1i} and e_{2i} of M is 1, the pair $a_i\,a_i'$ separates one but not both of the pairs $a_1\,a_1'$ and $a_2\,a_2'$, and hence has one side on each of the parts of the polygon π between a_1' and a_2'; hence e_{2i} of M_1 is 1.
(c) The element e_{ij} of M_1, where $i, j \neq 1, 2$, is the same as the element e_{ij} of M, for it is obvious that the above transformation does not affect the mutual relations of two pairs neither of which is $a_1\,a_1'$ or $a_2\,a_2'$.

In the case where a_1 and a_1' have the same sense, it follows similarly that the matrix M_1 is obtained from the matrix M by adding the first row and column to the second row and column respectively and reducing each element modulo 2.

24. We shall next see that any transformation of the polygon by a single cut may be obtained as the resultant of a series of cuts of the simple kind just considered. First it is obvious that the polygon obtained by two cuts $\bar{a}_1 \equiv a_1 + a_2$ and $\bar{a}_1 \equiv \bar{a}_1 + a_3$ is the same as the polygon obtained by the cut $\bar{a}_1 \equiv a_1 + a_2 + a_3$, where for the first cut the parts of the polygon are joined along the sides a_1 and a_1', for the second along a_1 and $\bar{a}_1'$, and for the third along a_1 and a_1'. This shows that any transformation $\bar{a}_1 \equiv \Sigma_i a_i$ where the two parts of the polygon are joined along two sides a_1 and a_1', one of which has an end in common with $\bar{a}_1$, may be obtained by a series of transformations of the type $\bar{a}_1 \equiv a_1 + a_2$. In the case where the two parts of the polygon are joined along a pair of sides neither of which has an end in common with $\bar{a}_1$, we note that such a transformation may be obtained as the resultant of two transformations of the preceding type.* Thus any transformation of the polygon by a single cut may be accomplished by a series of transformations of the type $\bar{a}_1 \equiv a_1 + a_2$, and consequently any transformation of the polygon by the method of cutting may be accomplished by a series of transformations of the same type.

25. In § 23 we saw that M_1 can be obtained from M by adding the first row to the second row, adding the first column to the second column, and reducing each element modulo 2. From the theory of matrices† it follows that the ith row of M may be added to the jth row and the ith column to the jth column by multiplying M on the left by a certain matrix A of determinant 1 and multiplying the result on the right by the conjugate matrix A'. Since by § 24 any transformation by the method

* For example, the result of the cut $\bar{a}_k \equiv a_1 + \cdots + a_j + a_k + a_l + \cdots + a_m$, where the two parts are joined along a_k and a_k', is the same as the result of the cut $\bar{a}_k \equiv a_1 + \cdots + a_j + a_k$ followed by the cut $\bar{a}_k \equiv \bar{a}_k + a_l + \cdots + a_m$, where the parts are joined along a_k and a_k' in the first case and along $\bar{a}_k$ and $\bar{a}_k'$ in the second.

† See Veblen and Franklin, these Annals, vol. 23, pp. 1–15.

of cutting may be effected by a series of cuts of the type described in § 23, it follows that *if the polygon π_1 is obtained from the polygon π by the method of cutting, the separation matrix M_1 of π_1 may be obtained from the separation matrix M of π by multiplying M on the left by a matrix A of determinant* 1 *and on the right by the conjugate matrix A', and then reducing each element modulo* 2.

The converse of this theorem is not true; we shall return to this question in a later paragraph.

26. Let us consider a polygon to which Reduction 2 of § 7 may be applied. In the separation matrix of the polygon the row and column which correspond to the conjugate pair $a_i\, a_i'$ will be made up wholly of zeros. The separation matrix of the polygon that is obtained by carrying out Reduction 2 is the matrix obtained by striking out the row and column of zeros. Reduction 1 is an operation of the type considered in § 25. Hence, *the connectivity of the manifold is one greater than the rank of the separation matrix of the polygon.*

27. **The Normalization of the Separation Matrix.** We have seen that a polygon whose conjugate pairs of sides correspond to the circuits of a fundamental set may be reduced to normal form by the method of cutting without reducing the number of sides. The separation matrix of the normalized polygon of a two-sided manifold is a matrix in which $e_{2n-1,\,2n}$ and $e_{2n,\,2n-1}$ $(n = 1, 2, \cdots, (R_1 - 1)/2)$ are equal to 1 and every other element is 0; the separation matrix of the normalized polygon of a one-sided manifold is a matrix in which $e_{n,\,n}$ $(n = 1, 2, \cdots, (R_1 - 1))$ is 1 and every other element is 0. These matrices are normal forms for symmetric matrices (mod 2) of determinant 1.* As a result of these considerations and § 25 we have the theorem: *If M is the separation matrix of a polygon whose vertices constitute a single conjugate set, there exists a matrix A of determinant* 1 *such that the product $A\,M\,A'$ is equivalent modulo* 2 *to the normal form of a symmetric matrix of determinant* 1, *and such that A corresponds to a series of cuts on the polygon.*

28. We have seen that, when the polygon is in normal form, the separation matrix is also in normal form. The converse of this statement is, however, not true, as we saw in § 21. Instead we have the following theorem: *If the separation matrix is in normal form, the polygon may be normalized by a series of cuts of which the corresponding matrix A is the identity, modulo* 2.

Let us consider the one-sided and two-sided cases separately. In the one-sided case the polygon is in normal form or else there is a pair $a_i\, a_i'$ such that one of the two parts of the boundary between a_i and a_i' is

* See Veblen and Franklin, loc. cit., p. 14.

made up of the sides $a_j\, a_j'\, a_k\, a_k' \cdots a_l\, a_l'$ in that order. The cut joining the forward ends of a_i and a_i' gives the following transformation on the circuits of the fundamental set when the two parts are joined along a_i and a_i':

$$
\begin{aligned}
\bar{a}_1^1 &\equiv a_1^1 \\
&\cdot \\
&\cdot \\
&\cdot \\
\bar{a}_i^1 &\equiv a_i^1 + 2a_j^1 + 2a_k^1 + \cdots + 2a_l^1 \\
&\cdot \\
&\cdot \\
&\cdot \\
\bar{a}\mu^1 &\equiv a\mu^1.
\end{aligned}
$$

The matrix A corresponding to this cut has a main diagonal made up of 1's and no other elements excepting 0's and 2's. This transformation has increased by one the number of pairs of sides which are in the order $a_i\, a_i'$ on the boundary of the polygon. By repeating this process the polygon may be reduced to normal form.

In the two-sided case, if the polygon is not in normal form, there must be some group of four sides $a_i\, b_i\, a_i'\, b_i'$ such that between two elements of the group, say between b_i and a_i', there are one or more groups of four consecutive sides $a_j\, b_j\, a_j'\, b_j'$. A cut joining the forward ends of b_i and b_i' gives a matrix A which is equal mod 2 to the identity, and so does a cut joining the forward ends of the sides $\bar{a}_i$ and $\bar{a}_i'$ obtained from the first cut. This transformation increases the number of groups of four consecutive sides of the form $a_j\, b_j\, a_j'\, b_j'$ and may be continued until the polygon is normalized.

29. From these theorems we can now deduce an important theorem analogous to the theorem given by Poincaré on page 70 of the Fifth Complement. *Given two fundamental sets $a_1, a_2, \cdots, a_\mu$ and $b_1, b_2, \cdots, b_\mu$; in order that there shall exist a fundamental set $c_1, c_2, \cdots, c_\mu$, into which the a's are transformable by a homeomorphism of the manifold with itself and which are homologous (mod 2) with $b_1, b_2, \cdots, b_\mu$ respectively, it is necessary and sufficient that the separation matrix of the a's shall be the same as that of the b's.*

If the a's are transformable into the c's by a homeomorphism of the manifold, this homeomorphism determines a homeomorphism of the polygon of the a's with that of the c's. Hence the separation matrix M_a of the a's is the same as the separation matrix M_c of the c's. By § 14 it is possible to pass from the c's to the b's by the method of cutting. This determines a set of homologies connecting the c's with the b's, and

if A is the matrix of this set of homologies, we have by § 25,

$$M_b = A' \cdot M_c \cdot A,$$

where M_b is the separation matrix of the b's. By hypothesis we have a set of homologies, $b_i \sim c_i$ (mod 2). But there cannot be more than one set of homologies (mod 2) connecting the b's and the c's, since otherwise there would be homologies of the form $c_i \sim c_j$ (mod 2) among the c's. Hence A is the identity matrix and $M_b = M_c$. Hence $M_a = M_b$.

Conversely, let us suppose that $M_a = M_b$. The a's and the b's respectively can be converted by the method of cutting into fundamental sets $d_1, d_2, \cdots, d_\mu$ and $f_1, f_2, \cdots, f_\mu$ respectively whose polygons are in normal form. Then, if a sequence of cuts is applied to $f_1, f_2, \cdots, f_\mu$ which is homeomorphic with a sequence of cuts which converts $d_1, d_2, \cdots, d_\mu$ back into $a_1, a_2, \cdots, a_\mu$, the f's are evidently converted into a fundamental set $c_1, c_2, \cdots, c_\mu$ which is capable of being transformed into $a_1, a_2, \cdots, a_\mu$ by a homeomorphism of the manifold with itself. Hence $M_a = M_c$ and therefore $M_c = M_b$. But the c's have been obtained from the b's by the method of cutting and so are related to them by an equation of the form $A' \cdot M_c \cdot A = M_b$. By § 28 the c's can be obtained from the b's by a series of cuts for which A is the identity. Since there cannot be more than one set of homologies (mod 2) relating the b's and the c's, it follows that

$$\begin{array}{l} b_1 \sim c_1 \\ b_2 \sim c_2 \\ \;\cdot \quad\;\; \cdot \\ \;\cdot \quad\;\; \cdot \\ \;\cdot \quad\;\; \cdot \\ b_\mu \sim c_\mu \end{array} \quad \text{(mod 2)}.$$

30. **The Matrix of Signed Separations.** In the case of the two-sided manifold we may give an algebraic sign to the separations of pairs of sides of the polygon. First let us assign a sense arbitrarily to the boundary of the polygon. Of each conjugate pair one side agrees in sense with the boundary and the other disagrees with it; let the side which agrees in sense with the boundary be designated by a_i, and the other by a_i'. Suppose an arc drawn joining the forward ends of a_i and a_i', and let α be the part of the polygon on whose boundary the two sides a_i and a_i' appear. If $a_j\, a_j'$ separates $a_i\, a_i'$ and the side a_j is on the boundary of β, we will say that $a_j\, a_j'$ separates $a_i\, a_i'$ *positively*; if a_j is on α, we will say that $a_j\, a_j'$ separates $a_i\, a_i'$ *negatively*.

As an immediate consequence of the above definitions it follows that *if $a_j\, a_j'$ separates $a_i\, a_i'$ positively, then $a_i\, a_i'$ separates $a_j\, a_j'$ negatively*. In like manner it follows that *reversing the senses of the sides a_i and a_i' changes the sign of every separation by that pair.*

Let us give each non-zero element of the separation matrix the sign $+$ or $-$ according as it stands for a positive or a negative separation. The resulting matrix will be called the *matrix of signed separations.* From the last paragraph it follows that this matrix is skew-symmetric.

31. Consider a cut $\bar{a}_1 \equiv a_1 + a_2$ (cf. Fig. 2). The 1-cell $\bar{a}_1$ divides the polygon into two parts one of which has on its boundary $\bar{a}_1$, a_1, and a_2. If $\bar{a}_1$ is given a sense which disagrees with the sense of a_1 on the boundary of this part, then the signed separations by $\bar{a}_1\,\bar{a}_1'$ on π_1 will be identical with the signed separations by $a_1\,a_1'$ on π. With this convention in assigning a sense to $\bar{a}_i$, we will prove that *if π_1 is obtained from π by a cut $\bar{a}_1 \equiv a_1 + a_2$, the matrix of signed separations S_1 of π_1 may be obtained from the matrix of signed separations S of π by multiplying the first row by -1 and adding it to the second row, and performing the same operation on columns.*

Proof: The separation matrix of any polygon can be obtained from the matrix of signed separations by reducing each element of the latter modulo 2. The matrix given by the theorem when each element is reduced modulo 2 is the separation matrix of the transformed polygon. (Cf. § 25.) Therefore the proof of the theorem reduces to the proof of the facts (1) that the matrix of the transformed polygon given by the theorem contains no element different from 0, 1, and -1, and (2) that by the method given in the theorem the proper sign is attached to each element. To prove (1) it is sufficient to show that if e_{1i} and e_{2i} are both different from 0 they have the same sign. This means that if $a_i\,a_i'$ separates both $a_1\,a_1'$ and $a_2\,a_2'$, it separates both positively or both negatively, which follows from the fact that a_1 and a_2 have the same sense. To prove (2) consider first the case where $a_i\,a_i'$ separates $a_1\,a_1'$ but does not separate $a_2\,a_2'$ on π. We are to show that e_{2i} of S_1 is 1 or -1 according as e_{1i} of S is -1 or 1. This follows from the fact that a_i or a_i' is on the part of the boundary of π between a_1 and a_1' which does not contain a_2'. Finally consider the case where $a_i\,a_i'$ separates $a_2\,a_2'$ but does not separate $a_1\,a_1'$. In this case the transformation does not affect the separation of $a_2\,a_2'$ by $a_i\,a_i'$, which gives that if e_{1i} of S is 0, e_{2i} of S_1 is the same as e_{2i} of S.

32. Consider the cut $\bar{a}_1 \equiv a_1 + a_2'$. This can be reduced to the case treated in § 31 by changing the sense of each of the two sides a_2 and a_2'. This changes the sign of each element in the second row and each element in the second column (§ 30). Now carry out the transformation $\bar{a}_1 \equiv a_1 + a_2$; the corresponding transformation on S multiplies the first row and column by -1 and adds them to the second row and column respectively. Finally reverse the senses of a_2 and a_2' again and carry out the corresponding change on the matrix. The result may be expressed

as follows: *If π_1 is obtained from π by a cut $\bar{a}_1 \equiv a_1 + a_2'$, the matrix S_1 of signed separations of π_1 may be obtained from the matrix S of signed separations of π by adding the first row to the second row and performing the same operation on columns.*

33. By omitting the phrase "modulo 2" in the theorems of § 25 and § 27 and replacing M and A by S and B respectively, we get two theorems concerning the matrix of signed separations. That these theorems are true follows easily from §§ 31, 32. Corresponding to the theorem of § 28 we have: *If the matrix of signed separations is in normal form, the polygon may be normalized by a set of cuts of which the matrix B is the identity.*

To prove the theorem we need only (cf. § 28) show that the matrix B corresponding to the cut $\bar{b}_i \equiv b_i + a_j + b_j + a_j' + b_j'$ is the identity. This cut may be effected by the following series of cuts:

$$x_1 \equiv b_i + a_j, \qquad x_2 \equiv x_1 + b_j, \qquad x_3 \equiv x_2 + a_j', \qquad \bar{b}_i \equiv x_3 + b_j'.$$

The product of the matrices of these transformations is the identity.

By proceeding as in § 29 we may now establish a theorem identical with that of § 29 with omission of the modulo 2 condition. This is equivalent to the theorem given by Poincaré (l.c., p. 70).

34. Given any series of cuts on the polygon we have seen that there corresponds to it a matrix B whose determinant is 1. As a result of the first theorem of § 33 we have that there exists more than one series of cuts corresponding to a given matrix, if there exists one. It can be shown however, by means of a simple example, that *not every matrix of determinant* 1 *corresponds to the transformation of a given polygon by a series of cuts.*

35. **Criterion for a Non-singular Circuit.** *Any simple circuit which is not homologous to zero is homologous to a linear combination, with coefficients relatively prime, of circuits of any fundamental set.**

Proof: The circuit may be deformed into one which passes through the point A of any fundamental set F. The image on π of the circuit will be a set of non-intersecting arcs. By the method of cutting we may obtain a polygon π_1 on which the image of the circuit is an arc joining two consecutive vertices.

The separation matrix M_1 of π_1 is equal to AMA' (modulo 2) where M is the separation matrix of π and A is the product of a set of matrices $A_k A_j A_i \cdots A_1$, each of which corresponds to a single cut and is therefore of determinant 1. The matrix $A_k' A_j' A_i' \cdots A_1'$ is the matrix of the homology transformation of the circuits of F into the circuits of F_1. (See § 22.) The elements of the ith row of this matrix are the coefficients of a combination of the circuits of F which is homologous to the circuit

* Poincaré proves this theorem and its converse for two-sided manifolds, l.c., page 70.

C_i' of F_1, $C_i' \sim \Sigma_j e_{ij} C_j$. Since the matrix is of determinant 1 the theorem follows.

From the foregoing it is evident that the theorem just proved is true in the case of a two-sided manifold without the restriction in the hypothesis to circuits which are not homologous to zero. It is equally evident that the restriction is necessary in the case of a one-sided manifold, for a circuit whose image on the polygon together with two sides C_i and C_i' which have the same sense bounds a part α of the polygon is equivalent to $2C_i$. However C_i, a circuit of the fundamental set, is homologous to a linear combination with coefficients relatively prime of circuits of any fundamental set. Thus we have the result that on a one-sided manifold any simple circuit is homologous to a linear combination with coefficients relatively prime of any fundamental set, or else it is homologous to a linear combination with coefficients containing 2 as a highest common factor. This factor 2 is the *coefficient of torsion* of a one-sided manifold.

Any linear combination with coefficients relatively prime of circuits of a fundamental set for a two-sided manifold is homologous to a simple circuit.

Proof: The method of proof will be to show that a matrix B with an arbitrary first row, provided the elements are relatively prime, may be built up by taking the product of a set of matrices each of which corresponds to a cut on the polygon. First reduce the polygon to normal form. Let D be the matrix to which this reduction corresponds. We shall now find a matrix C such that $B = C \cdot D$ has an arbitrary first row and such that the matrix C corresponds to a set of cuts. That B may have an arbitrary first row, it is sufficient that the first row of C may be chosen arbitrarily.

The two transformations which follow can be carried out on the normalized polygon and each transformation leaves the polygon in normal form.

$$(1) \qquad \bar{a}_{2n-1} \sim a_{2n-1} + b_{2n},$$

$$(2) \qquad \bar{a}_{2n-1} \sim a_{2n-1} + a_{2m-1} \qquad \text{followed by} \qquad \bar{b}_{2m} \sim b_{2m} - b_{2n}.$$

The matrix B_1 corresponding to transformation (1) is (for $n = 2$) of the form:

$$B_1 = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{vmatrix}.$$

The matrix corresponding to the transformation (2) is (for $n = 2$, $m = 1$) of the form:

$$B_2 = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{vmatrix}.$$

By taking products of matrices of the type B we may obtain a matrix of the form:

$$B_3 = \begin{vmatrix} a_{11} & a_{12} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & a_{43} & a_{44} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{55} & a_{56} & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{65} & a_{66} & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & a_{2p-1,\,2p-1} & a_{2p-1,\,2p} \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & a_{2p,\,2p-1} & a_{2p,\,2p} \end{vmatrix},$$

where $a_{i,\,i+1}$ and $a_{i,\,i}$ are any two integers relatively prime and where

$$\begin{vmatrix} a_{i,\,i} & a_{i,\,i+1} \\ a_{i+1,\,i} & a_{i+1,\,i+1} \end{vmatrix} = 1.$$

This follows from the fact that any two-rowed matrix of determinant 1 may be normalized by elementary transformations on the rows alone. The elements $a_{i,\,i}$ and $a_{i,\,i+1}$ may be chosen so that $e_{1,\,i}/a_{i,\,i} = e_{1,\,i+1}/a_{i,\,i+1}$, where $e_{1,\,i}$ and $e_{1,\,i+1}$ are elements of the arbitrarily given first row of C. Then, by application of matrices of type B_2 above, any odd row of B_3 may be added to the first row a sufficient number of times to give the arbitrary first row of C. This completes the proof of the theorem.

36. **Intersections of Circuits of Fundamental Set.** Consider two sensed circuits C_1 and C_2 on a two-sided manifold. Let them have a point P in common. A 2-cell a_1^2 may be constructed which contains P, and no other point common to the two circuits, as an interior point and which contains a simple arc of each of the circuits on the interior. Let one of the senses of description of the boundary be designated as positive. Let the forward end of the arc of C_i which is interior to a_1^2 be called a_{i1}^0 and the other end a_{i2}^0. If the points $a_{11}^0\, a_{12}^0$ separate the points $a_{21}^0\, a_{22}^0$, the

two circuits C_1 and C_2 will be said to *intersect* at P. If the two circuits intersect at P and the point $a_{22}{}^0$ is on the part of the boundary of $a_1{}^2$ that runs positively from $a_{11}{}^0$ to $a_{12}{}^0$, C_2 will be said to intersect C_1 *positively*: if $a_{21}{}^0$ is on that arc C_2, it will be said to intersect C_1 *negatively*. We have as an obvious theorem: *If C_2 intersects C_1 positively at the point P, then C_1 intersects C_2 negatively at the point P.*

37. Consider now two circuits C_1 and C_2 which have more than one point in common. A 2-cell may be constructed at each common point as in § 36. These 2-cells may be assigned senses in such a way that they all agree in sense. Making use of these sensed 2-cells, we may determine the number of positive and the number of negative intersections of the circuit C_2 with the circuit C_1. Let $N(C_2, C_1)$ be a positive or a negative number equal to the number of positive intersections of C_2 with C_1 minus the number of negative intersections of C_2 with C_1. As a result of this definition and the theorem of § 36 we have

$$N(C_2, C_1) = -N(C_1, C_2).$$

The following theorems may be easily proved:

If $C_1 \sim 0$, and C_2 is any circuit whatever, then $N(C_2, C_1) = 0$.

If $C_3 \equiv C_1 + C_2$, and C_4 is any circuit whatever, then

$$N(C_4, C_3) = N(C_4, C_1) + N(C_4, C_2).$$

If $C_1 \sim C_2$, and C_3 is any circuit whatever, then $N(C_3, C_1) = N(C_3, C_2)$.

38. **The Intersection Matrix.** Let us consider the intersections of pairs of circuits of a fundamental set, and let us construct a matrix of $2p$ rows and $2p$ columns by making the element e_{ij} equal the number $N(C_j, C_i)$. Since the circuits are simple circuits and no two have more than one point in common, the elements of the matrix will be 0, 1, and -1. Every element e_{ii} will be 0; the element $e_{j,i}$ will be the negative of the element e_{ij}. Thus the matrix is skew-symmetric.

39. A cut $\bar{a}_1 \equiv a_1 + a_2$ performs a certain transformation on the circuits of the fundamental set. According to § 37 the intersections by the circuit on which $\bar{a}_1$ is imaged are obtained by adding the rows corresponding to C_1 and C_2 in the intersection matrix N. Then to get the intersection matrix of the transformed fundamental set we add the second row of N to the first row and perform the same operation on columns. This may be accomplished by multiplying N on the left by the matrix which is the inverse of the conjugate of the matrix B used in § 33, and by multiplying on the right by the conjugate matrix.

From this it follows at once that *if the fundamental set F_1 is obtained from the fundamental set F by the method of cutting, the intersection matrix*

N_1 of F_1 and the intersection matrix N of F satisfy the relation $N_1 = T \cdot N \cdot T'$, *where T is a matrix of determinant* 1.

40. The polygon was normalized by the method of cutting. When the polygon is in normal form, the intersection matrix of the corresponding fundamental set is in normal form, as can be seen by constructing a neighborhood of the point A in the manner of § 36, and the matrix of signed separations is also in normal form. These two normal forms are the same. The matrix of signed separations of the original polygon is normalized by a matrix $B = (B_k \cdot \cdots \cdot B_2 \cdot B_1)$; the intersection matrix is normalized by a matrix $(B_k')^{-1} \cdots (B_2')^{-1}(B_1')^{-1}$. From this it follows by a simple computation with the matrices that *the intersection matrix is the negative of the reciprocal of the matrix of signed separations.*

TWO GENERALIZATIONS OF THE STIELTJES INTEGRAL.

By P. J. Daniell.

1. **Introduction.** In a paper on a General Form of Integral* the author gives an example of an integral with respect to a function which is not of limited variation, namely,

$$\int_0^1 f(x)d \log x,$$

which can be defined when $f(x)/x$ is continuous $(0 \leqq x \leqq 1)$. The first part of the present paper is an extension to a general class of integrals of this type. In other words, it considers integrals with respect to a general function $\alpha(x)$ which can be defined when appropriate restrictions are laid on the integrand $f(x)$.

The Stieltjes integral differs from the more usual integral in that it is invariant under a transformation of the independent variable which leaves relative position unchanged if the mass-distribution is transformed in the corresponding manner. It is less dependent on metrical geometry. This suggests an extension of the concept to an integral which is an operation on sets directly without any interpolation of measure. This concept may be useful in the theory of sets of points, but apart from that it opens an interesting field.

2. **Integration with respect to any function.** The notion we are about to develop can be extended to several dimensions but we prefer to give the development only for a single variable. Let $\alpha(x)$ be some function defined for all real values of x. If it is not, we may extend its definition by assigning to it the value 0 wherever it is not defined.

Relative to $\alpha(x)$ a point x is said to be proper on the right if an interval xx' can be found of which x is the left-hand endpoint and in which $\alpha(x)$ is of limited variation. It is improper on the right if such an interval cannot be found. Similarly it is proper (improper) on the left if $\alpha(x)$ is of limited variation in some (no) interval $x''x$ to the left of x. Every point is either proper or improper on the right and also either proper or improper on the left. If a point is proper on both sides, it is said to be "proper" (without qualification), while if it is improper on either side, it is said to be "improper." If further it is improper on both sides, we may say that it is "completely improper."

Theorem. *If p_1, p_2, $\cdots$ is an increasing sequence of improper points and if the limit of the sequence is p, p is at least improper on the left.*

* P. J. Daniell, these Annals, vol. 19 (1918), p. 279.

For if not, an interval $p'p$ could be found to the left of p such that in it α is of limited variation. But such an interval would enclose a point p_n of the sequence so that α is of limited variation in both $p'p_n$ and p_np. This contradicts the hypothesis that each p_n is improper. Similarly if a point p is approached by a sequence of improper points from the right, p is improper on the right.

It follows that the set of improper points is *closed* in the sense of the theory of sets (e.g., the set of all non-negative numbers is closed but not compact). Again any proper point lies strictly within an interval of proper points. Hence the set of proper points consists of a countable (that is, zero, finite, or denumerably infinite) set of non-compact intervals, δ_n, the complement of the closed set of improper points, K. A non-compact interval is usually an open interval but it is an open question whether the interval consisting of all real numbers should be called an open interval or not. As a set of points it is closed since it includes its derived set.

Every point which is improper on the left only is the left-hand endpoint of some interval δ_n, for otherwise it would be the limit from the right of a sequence of improper points; and similarly for points improper to the right only. Hence the set of points which are improper on one side only is countable. A point of K which is not an endpoint of an interval δ is completely improper. But an endpoint may also be completely improper. For example, let $\alpha(x)$ be defined as

$$\begin{aligned}\alpha(x) &= \sin 1/x \qquad & x &\neq 0,\\ &= 0 & x &= 0.\end{aligned}$$

Then the proper intervals are $(-\infty < x < 0)$, $(0 < x < +\infty)$. $x = 0$ is the endpoint of two intervals δ, but it is completely improper. On the other hand if

$$\begin{aligned}\alpha(x) &= \sin 1/x \qquad & x &> 0,\\ &= 0 & x &\leqq 0,\end{aligned}$$

the proper intervals are as before but $x = 0$ is improper to the right only.

Let Δ be a closed and compact interval contained strictly within an interval δ_n. Then $\alpha(x)$ must be of limited variation on Δ. For every point of Δ is strictly within an interval in which α is of limited variation. By the Heine-Borel theorem a finite number of these intervals can be found covering Δ completely between them so that $\alpha(x)$ is of limited variation over their sum, which includes Δ.

It follows that there can only be a countable number of points in any Δ (and therefore in $\Sigma\delta$) at which $\alpha(x)$ is discontinuous. Let $\Delta = (x_1, x_2)$ be an interval enclosed strictly within a δ_n and such that $\alpha(x)$ is continuous

at x_1, x_2. We define the "mass" of Δ as

$$m(\Delta) = \alpha(x_2) - \alpha(x_1).$$

Let $\varphi_n(x)$ be a function equal to 1 on such an interval Δ_n and 0 elsewhere. We define

$$\int \varphi_n(x) d\alpha(x) = m(\Delta_n).$$

If $f(x)$ is a linear combination of a finite number of such functions $\varphi_n(x)$,

$$f(x) = c_1\varphi_1(x) + \cdots + c_n\varphi_n(x),$$

we define

$$\int f(x) d\alpha(x) = c_1 m(\Delta_1) + \cdots + c_n m(\Delta_n).$$

The definition of the general integral given in the paper referred to above depends on a class T_0 of functions for which the integrals are supposed to be already defined.

We specify this class T_0 to be the class of functions just mentioned, linear combinations of functions of type φ. Evidently a multiple of the modulus of a function, and the sum of two functions of this class is of the same class. Hence T_0 satisfies the required conditions. For such functions

$$(C) \qquad \int cf(x) d\alpha(x) = c\int f(x) d\alpha(x),$$

$$(A) \qquad \int (f_1 + f_2) d\alpha = \int f_1 d\alpha + \int f_2 d\alpha,$$

$$(M) \qquad \left|\int f d\alpha\right| \leqq \max |f| \times \Sigma_n \text{ (variation of } \alpha \text{ on } \Delta_n).$$

It is only necessary to prove that postulate (L) is satisfied. This states that if $f_1, f_2, \cdots$ is a non-increasing sequence of functions of class T_0 which approaches 0 everywhere as a limit, then

$$\lim \int f_n d\alpha = 0.$$

Since $f_n \leqq f_1$, and f_1 differs from 0 only over a finite set of intervals $\Delta_1, \Delta_2, \cdots, \Delta_k$ each contained within a δ of CK,

$$\int f_n d\alpha = \int_{\Delta_1} f_n d\alpha + \int_{\Delta_2} f_n d\alpha + \cdots + \int_{\Delta_k} f_n d\alpha.$$

But in each Δ_i $(i = 1, 2, \cdots, k)$, $\alpha(x)$ is of limited variation so that, by the classical theory of the Stieltjes integral,

$$\lim \int_{\Delta_i} f_n d\alpha = 0.$$

Therefore all the required conditions for the definition of the integral are satisfied, and the definition can be extended, as in the paper to which we have referred, so as to include all integrands, $f(x)$, summable with respect to $\alpha(x)$.

3. **Standard summable function.** Every interval Δ is contained strictly within a δ of CK, and therefore every corresponding φ is 0 at points of K, and approaches 0 from either side as a point of K is approached.

A function of class T_1 is the limit of a non-decreasing sequence of functions of class T_0 and a summable function is less than a function of class T_1 and greater than the negative of such a function. Consequently all summable functions must, by definition, vanish at every point of K. But we can prove more than this. Suppose that the endpoint of an interval δ is improper in the direction in which the interval lies. It will be shown that 0 is a sublimit of any summable function $f(x)$ as x approaches the endpoint along the interval. Suppose that the point $x = b$ is the right-hand endpoint of an interval δ, and that b is improper on the left. Let a be a point within δ at which α is continuous, and let $\omega(x)$ be the variation of $\alpha(x)$ between a and x. Since b is improper on the left, $\omega(x)$ is unbounded as x approaches b. If f is summable with respect to α, it is also summable with respect to ω and its modulus is also summable. Thence

$$\lim \int_a^x f(x)d\omega(x) \qquad (x \doteq b)$$

must exist, where $f(x) \geqq 0$. If no sublimit of f is 0, the lower limit of $f(x)$ must be positive. Let it be $l > 0$. Then an interval c, b can be found within which $f(x) > l/2$. Hence

$$\int_a^x f(x)d\omega(x) \geqq l/2[\omega(x) - \omega(c)],$$

which increases without limit as x approaches b.

If $\alpha(x)$ is sufficiently irregular, K may consist of every point, and in this case the only summable function will be that which is identically 0, so that for some functions $\alpha(x)$ the definition of the integral will be valueless. But in many problems although $\alpha(x)$ is not of limited variation when the whole interval is taken, it is of limited variation when a set is eliminated by a covering set of intervals.

Now although every summable function must vanish at every point of K, a function can be found which is non-negative and summable, which vanishes nowhere except on points of K, and which has 0 for a sublimit only as x approaches an endpoint which is improper on the δ-interval side. Such a function plays the part of the function $h(x) = 1$ in the case of the ordinary Stieltjes integral.* Within each interval δ_n of the set CK choose a point P_n $(x = x_n)$ at which $\alpha(x)$ is continuous. Using P_n as a base we can define a variation function $\omega_n(x)$ for every point x in δ_n, which is 0 at P_n and non-decreasing as we proceed from P_n in either direction.

* Cf. P. J. Daniell, these Annals, vol. 21 (1920), p. 203.

If x belongs to δ_n, define

$$\begin{aligned}\beta(x) &= \omega_n(x+0) \qquad & x > x_n,\\ &= \omega_n(x-0) \qquad & x < x_n,\\ &= 0 \qquad & x = x_n.\end{aligned}$$

Since $\alpha(x)$ is continuous at $x = x_n$,

$$\omega_n(x+0) = \omega_n(x-0) = 0.$$

Define

$$\begin{aligned}h(x) &= 0 \text{ if } x \text{ belongs to } K,\\ &= 1/[2^{2n} + \beta^2(x)] \text{ if } x \text{ belongs to } \delta_n.\end{aligned}$$

We assert that $h(x)$ is summable with respect to $\alpha(x)$. It is evidently non-negative and differs from 0 except on K. It is also the limit of a sequence of functions of class T_0. Let Δ_n be an interval of length d of which x_n is the left-hand endpoint contained within a δ_n, and such that α is continuous at $x_n + d$ (we have already chosen x_n to be a point of continuity of α). Let

$$\begin{aligned}h_n(x) &= h(x) \text{ on } \Delta_n,\\ &= 0 \text{ otherwise.}\end{aligned}$$

Then $h_n(x)$ is summable with respect to α and

$$\int h_n(x)d\alpha(x) = \int_{\Delta_n} h(x)d\alpha(x).$$

The latter is almost an ordinary Stieltjes integral, for on Δ_n $\alpha(x)$ is of limited variation and $h(x)$, although not continuous, is monotone and bounded.

$$\begin{aligned}\int h_n(x)\,|d\alpha(x)| &\leqq \int_{x_n}^{x_n+d} h(x)d\omega(x)\\ &= \int_{x_n}^{x_n+d} h(x)d\beta(x).\end{aligned}$$

Here $\int h\,|d\alpha|$ denotes the modular integral of h corresponding to $\int h d\alpha$, and the second inequality follows from the fact that at every point of continuity of α, ω and β coincide in value. Let us now make the Lebesgue transformation,*

$$\beta(x) = t.$$

At a discontinuity of α, and therefore of β, there is an interval of values of t corresponding to the one value of x, but over this discontinuity x,

$$\begin{aligned}\int_x h d\beta &= \frac{\beta(x+0) - \beta(x-0)}{2^{2n} + \beta^2(x)}\\ &= \frac{\beta(x+0) - \beta(x-0)}{2^{2n} + \beta^2(x+0)}\\ &< \int_{\beta(x-0)}^{\beta(x+0)} \frac{dt}{2^{2n} + t^2}.\end{aligned}$$

* H. Lebesgue, Comptes Rendus, vol. 150, p. 86.

Therefore, finally,

$$\int h_n(x)\,|d\alpha(x)| \leqq \int_0^{\beta(x_n+d)} \frac{dt}{2^{2n}+t^2}$$
$$< \pi/2^{n+1}.$$

Similarly for an interval Δ of which x_n is the right-hand endpoint,

$$\int h_n(x)\,|d\alpha(x)| < \pi/2^{n+1}.$$

If then $h'(x) = h(x)$ on each of a countable set of intervals Δ contained within the δ_n of CK, and 0 otherwise,

$$\int h'(x)\,|d\alpha(x)| < \pi(\tfrac{1}{2} + \tfrac{1}{4} + \cdots)$$
$$= \pi.$$

But h is the limit of a non-decreasing sequence of functions of type h' and therefore, by a theorem in the paper to which we referred at the beginning, h is summable with respect to α. Many other such functions can easily be constructed by the reader.

Now if $f(x)$ is any function, summable with respect to α, it must vanish wherever h vanishes, so that a function φ can be found such that

$$f(x) = \varphi(x)h(x).$$

If Δ is *any* interval, define its "mass" as

$$m_1(\Delta) = \int_\Delta h(x)d\alpha(x).$$

Then $m_1(\Delta)$ is an additive function of intervals, by means of which we can define the more usual type of integral (Radon-Young integral).

If ψ is a step-function (that is, constant over each of a finite number of sub-intervals), evidently

$$\int \psi(x)dm_1(e) = \int \psi(x)h(x)d\alpha(x),$$

where e denotes the variable set of integration. Hence, step by step, it can be proved that if ψ is summable with respect to m_1, ψh is summable with respect to α and vice versa, and that

$$\int \psi(x)dm_1(e) = \int \psi(x)h(x)d\alpha(x).$$

Finally, therefore, if $f(x)$ is summable with respect to $\alpha(x)$, it can be expressed in the form $\psi(x)h(x)$, where $\psi(x)$ is summable with respect to $m_1(e)$ and

$$\int f(x)d\alpha(x) = \int \psi(x)dm_1(e).$$

This transforms the general type of integral considered to one of the Radon-Young type (extension of the Stieltjes integral by the methods of Lebesgue).

In the particular case given in the paper on a General Form of Integral and mentioned in the introduction to the present paper,

$$\alpha(x) = \log x,$$

the set K consists of the point $x = 0$ only, and

$$h(x) = 1/[4 + (\log x)^2].$$

If $0 \leqq x \leqq 1$,

$$\frac{x}{4} \leqq \frac{1}{4 + (\log x)^2}.$$

Hence $x/4$ is also summable with respect to $\log x$ and x satisfies the requirements for an $h(x)$ (in the interval 0, 1). Here

$$\begin{aligned} m_1(\Delta) &= \textstyle\int_\Delta x d \log x \\ &= \text{length of } \Delta, \end{aligned}$$

and if $f(x)/x$ is continuous in the interval, $f(x)$ is summable with respect to $\log x$.

Another example is obtained in the following way: Let E be a perfect set contained in the interval $J = (0, 1)$. Define $\alpha(x)$ as equal to x at all points of the intervals making up the set $J - E$, and at all irrational points whatever, but equal to 0 otherwise (that is, at rational points not belonging to $J - E$). Then the proper intervals δ_n consist of the open intervals forming $J - E$, while the set K consists of the points $x \leqq 0$, the set E, and the points $x \geqq 1$. If $f_1(x)$ is summable in the usual sense on the interval J, and if

$$\begin{aligned} f(x) &= 0 \text{ on } K \\ &= f_1(x) \text{ on } J - E = \Sigma\delta_n, \end{aligned}$$

then f is summable with respect to α and

$$\int f(x) d\alpha(x) = \Sigma_n \int_{\delta_n} f_1(x) dx.$$

4. **Integration of sets.** According to the paper on a General Form of Integral to which we have already referred, an integral can be defined, or at least extended, by means of certain simple processes such as addition, taking the greater or less of two functions and taking the limit of a monotone sequence. These processes have their analogies in the theory of sets of points, or of more general classes. We recall briefly the main principles of such processes.

There is assumed to be given a fundamental set J of elements, p. In this set are contained all the sets considered. The complement of J is the null set θ containing no elements. E_1E_2, the *product* of E_1, E_2, is the set of elements belonging to both. It corresponds both to an alge-

braic product and to the "logical product" (the lesser of two numbers). $E_1 + E_2$, the *sum* of E_1, E_2, is the set of points belonging to either and corresponds to the logical sum (the greater of two numbers) while when E_1, E_2 have no point in common it also corresponds to an algebraic sum. A vital distinction between products and sums in the theory of sets and in algebra is to be noted. The addition (multiplication) of a collection of algebraic numbers is impossible unless the collection has a power not greater than that of a denumerable infinity, and even then an infinite series (product) may not converge. But the sum (product) of any number of sets contained in J consists of the elements belonging to any one (every one) of the sets and this sum (product) is contained in J. If every element of E_1 is an element of E_2, we say that $E_1 < E_2$ (in particular $E < E$) and then $E_2 - E_1$ is the set E_2CE_1, where CE_1 is the set complementary to E_1. Subtraction is a useful process in the theory of sets but a dangerous one. For example, it is not, in general, true that $(A - B) + C = (A + C) - B$. Again there are no fractional or negative sets.

If $E_1, E_2, \cdots$ is a sequence of sets, we can form the sets,

$$F_n = E_n + E_{n+1} + \cdots \qquad (n = 1, 2, \cdots).$$

$F_1, F_2, \cdots$ is a decreasing sequence of sets whose limit F is the "complete limit" (according to Borel) of the sequence $[E_n]$. The limit F is defined as $F = F_1F_2 \cdots$, the set of elements belonging to every F_n.

If an element belongs to a finite or zero number of the sets, E_n, it is not contained in some F_n and is therefore not in F. If an element belongs to an infinity of the sets E_n, it belongs to every F_n and therefore to F. Hence F is the set of elements belonging to an infinity of the sets E_n. F may be called the "upper limit" of the sequence E_n. Similarly the lower limit or "restricted limit" (Borel) G is

$$G = G_1 + G_2 + \cdots$$
$$G_n = E_nE_{n+1} \cdots.$$

G is the set of elements belonging to all but a finite number of the sets E_n. If $F = G$, the sequence $[E_n]$ is said to converge to the limit F. This occurs when every element which belongs to an infinity of the E_n belongs to all but a finite number of them.

5. **Set-functions.** Let s be a real number. If to each value of s there corresponds a set $F(s)$ of elements p, we say that $F(s)$ is a set-function of s. The first analogy with the Stieltjes integral which suggests itself is the integral of $F(s)$ with respect to $E(s)$, where $F(s)$ is a continuous set-function and $E(s)$ a set-function of limited variation. But such an analogy

is valueless. On the one hand, even if we define a modular difference of classes, the sum of any number of such modular differences is always contained in J and a "variation" would be *always* limited.

On the other hand, a *continuous set-function must be constant.* A set-function $F(s)$ is said to be continuous at $s = a$ if, whenever $\lim s_n = a$, $\lim F(s_n)$ exists and equals $F(a)$. Let $F(s)$ be the function assumed to be continuous for all values of s $(-\infty < s < +\infty)$ and let $S(p)$ be the set of *real numbers* s for which p is an element of $F(s)$. When $\lim s_n = s$, $\lim F(s_n) = F(s)$. Therefore if $s_1, s_2, \cdots$ belong to $S(p)$, p belongs to $F(s_n)$ for all n and therefore to $F(s)$. Thus $S(p)$ is a closed set of real numbers. But $CS(p)$ corresponds in the same way to $J - F(s)$, which is also continuous as a set-function, and $CS(p)$ must also be closed. But a set of real numbers and its complement cannot both be closed unless one is the set of all numbers, the other of none. In consequence an element p either belongs to $F(s)$ for all s or for no values of s, and $F(s)$ is a constant set, the same for all s.

It is necessary to proceed in a different manner, using the essential distinctions between numbers and sets. Let $E(s)$ be an increasing set-function (there is no distinction between increasing and non-decreasing since any set is less than—and greater than—itself in the sense of inclusion). Then if $s_1 < s_2$, $E(s_1) < E(s_2)$. If $s_1, s_2, \cdots$ approaches s from below, $E(s_n)$ is an increasing sequence of sets possessing a limit. Also this limit is unique and may be called $E(s - 0)$. Similarly $E(s + 0)$ can be defined.

Define

$$\delta E(s) = E(s + 0) - E(s - 0).$$

If $e(s)$ is any set-function of s, let

$$\sigma e(s) \quad (S)$$

denote the sum of the sets $e(s)$ for all values of s belonging to the collection specified by S. Then if $F(s)$ is *any* set-function and $E(s)$ an increasing set-function, we define

$$\int F(s)dE(s) = \sigma[F(s)\delta E(s)] \quad (-\infty < s < +\infty).$$

This "set-integral" possesses certain interesting properties. For example, if $F(s)$, $G(s)$ are two set-functions,

$$\int (F + G)dE = \int FdE + \int GdE.$$

For, omitting the variable s,

$$\begin{aligned}\sigma[(F + G)\delta E] &= \sigma[F\delta E + G\delta E] \\ &= \sigma[F\delta E] + \sigma[G\delta E].\end{aligned}$$

Also

$$\int FGdE = (\int FdE)(\int GdE).$$

For if $s < t$,

$$\begin{aligned} \delta E(s) &= E(s+0) - E(s-0) \\ &< E(s+0) \\ &< E(t-0). \end{aligned}$$

Therefore $\delta E(s)$, $\delta E(t)$ have no element in common, and, by symmetry,

$$\begin{aligned} \delta E(s)\cdot\delta E(t) &= \delta E(s) \quad (s = t), \\ &= \theta \quad (s \neq t). \\ (\int FdE)(\int GdE) &= \sigma_s[F(s)\delta E(s)]\cdot\sigma_t[G(t)\delta E(t)] \\ &= \sigma_{s,\,t}[F(s)G(t)\delta E(s)\delta E(t)] \\ &= \sigma_s[F(s)G(s)\delta E(s)] \\ &= \int FGdE. \end{aligned}$$

If it is recalled that the product of a set into itself is equal to itself, the above inequality can be expressed in a form which reminds one of the Schwarz inequality in ordinary integration, this form being

$$(\int FGdE)^2 = (\int F^2dE)(\int G^2dE).$$

But in this theory of sets, infinite addition and multiplication can be as readily handled as finite processes, and by the same reasoning as before

$$\begin{aligned} \int(F_1 + F_2 + \cdots)dE &= \int F_1dE + \int F_2dE + \cdots, \\ \int F_1F_2\cdots dE &= (\int F_1dE)(\int F_2dE)\cdots. \end{aligned}$$

It follows from these equalities and the definition of a limit that

$$\int \lim F_n(s)dE(s) = \lim \int F_n(s)dE(s).$$

Let $r{*}t$ denote an interval $r < s \leqq t$, equal to the interval (r, t) (closed) with the point r omitted. Let

$$\begin{aligned} F_{r*t}(s) &= F(s) \quad (r < s \leqq t) \\ &= \theta \text{ otherwise.} \end{aligned}$$

We may define

$$\int_{r*}^{t} F(s)dE(s) = \int F_{r*t}(s)dE(s).$$

Then

$$\int_{r*}^{t} FdE + \int_{t*}^{u} FdE = \int_{r*}^{u} FdE.$$

Also

$$\begin{aligned} \int_{r*}^{t} JdE(s) &= \sigma(\delta E(s)) \quad (r < s \leqq t) \\ &= E(t+0) - E(r+0). \end{aligned}$$

For consider an element p. The numbers s can be placed in one of two classes $S(p)$, $CS(p)$ according to whether p belongs to $E(s)$ or not. Since $E(s)$ is increasing, any number in $CS(p)$ is less than any in $S(p)$ and, if

both classes exist, a "section" is obtained which defines a real number s (dependent on p). In this case p belongs to $E(s+0)$ but not to $E(s-0)$. Any element p must therefore satisfy one of three conditions, (a) p belongs to all $E(s)$, or (b) p belongs to no $E(s)$, or (c) p belongs to $\delta E(s)$ for some value of s. If p belongs to $E(t+0) - E(r+0)$, it satisfies neither (a) nor (b), and p belongs to $\delta E(s)$ for some s. *This s must lie in the interval, $r < s \leqq t$.* For if $s \leqq r$, p belongs to $E(s+0)$ which is excluded from $E(t+0) - E(r+0)$ in $E(r+0)$. Similarly if $s > t$, p is excluded from $E(s-0)$ which, however, includes $E(t+0)$. Then

$$E(t+0) - E(r+0) < \sigma[\delta E(s)] \quad (r < s \leqq t).$$

But if p belongs to some $\delta E(s)$ $(r < s \leqq t)$, it belongs to $E(s+0)$ and therefore to $E(t+0)$, while it does not belong to $E(s-0)$ and therefore not to $E(r+0)$. So that

$$\sigma[\delta E(s)] \quad (r < s \leqq t) < E(t+0) - E(r+0).$$

This proves the required equality. Evidently intervals (r, t), (r, t^*), (r^*, t^*) can be handled in the same manner.

6. **Directed continuity.** It has been proved that, if a set-function is continuous, it is constant and its properties are of little interest. But we can obtain a valuable class of functions if we restrict the continuity to be on one side only.

If for all values of s, the unique limit $F(s+0)$ exists and is equal to $F(s)$, then $F(s)$ is said to be continuous on the right. Similarly if $F(s-0) = F(s)$, $F(s)$ is continuous on the left. A set-function is called a step-function if it is constant over each of a finite number of intervals of s.

THEOREM. *A function which is continuous on the right is the limit of a sequence of step-functions.*

For if $F(s)$ is the given function and if

$$F_n(s) = F[t_n(s)],$$

where $2^n t_n(s)$ is the least integer not less than $2^n s$ (the integer equal to or just greater than $2^n s$), then

$$F(s) = \lim F_n(s).$$

Since $t_n(s)$ is a non-increasing sequence approaching s from above, if s is not a terminating fraction in the scale of 2,

$$\lim F_n(s) = F(s+0),$$

and otherwise, after some finite n, $t_n(s) = s$ so that

$$F_n(s) = F(s).$$

The theorem is thus proved. Now if $t_{i,n} = i2^{-n}$ where i is a positive or negative integer or zero,

$$\begin{aligned} \int F_n(s)dE(s) &= \Sigma_i F(t_{i,n})[\sigma\delta E(s)\ (t_{i-1,n} < s \leqq t_{i,n})] \\ &= \Sigma_i F(t_{i,n})[E(t_{i,n}+0) - E(t_{i-1,n}+0)]. \end{aligned}$$

Since

$$\int F dE = \lim \int F_n dE$$

the following important theorem is an immediate consequence:

THEOREM. *If for each value of s, $F(s)$, $E(s)$ are sets of points in one or more dimensions which are B-measurable (measurable in the sense of Borel); if $E(s)$ is an increasing set-function and $F(s)$ continuous on the right, then $\int F dE$ is also B-measurable.*

According to our primary definition for any F, an integral is obtained by an infinite process having the power of the continuum, but we see that, if F is continuous on the right, the integral can be obtained by passages from finite processes to the limit, processes which do not take the sets beyond the class of B-measurable sets.

The same result would hold if $F(s)$ were continuous on the left or if it were continuous on the left or right in each of a countable number of intervals, whose complementary set is a countable number of points ("countable" means zero, finite or denumerably infinite). The theorem also holds if "measurable in the sense of Lebesgue" is substituted for "B-measurable." This theorem has an immediate application to the theory of measurable functions. Let $E(s)$, $F(s)$ be the sets of points for which $e(p) < s$, $f(p) < s$, respectively, where $e(p)$, $f(p)$ are never-infinite functions of points p in one or more dimensions. If e, f are measurable (in either sense, this sense being retained throughout), so are$E(s)$, $F(s)$ measurable for each s. These sets are increasing and continuous on the left. For if $f(p) < s$, after some finite value of n, $f(p) < s - 2^{-n}$ and p belongs to $F(s - 2^{-n})$. If $f(p) \geqq s$, p belongs to no set $F(s - 2^{-n})$. The set $G(s)$ of points for which $e(p) + f(p) < s$ consists of the sum for all t of the sets where simultaneously $e(p) = t$, $f(p) < s - t$.

Now the set where $e(p) = t$ is the set $\delta E(t)$ and therefore

$$\begin{aligned} G(s) &= \sigma_t F(s-t)\delta E(t) \\ &= \int F(s-t)dE(t). \end{aligned}$$

Considered as a function of t, $F(s-t)$ is continuous on the right and, by our theorem, $G(s)$ is measurable and

$$g(p) = e(p) + f(p)$$

is measurable in the same sense as $e(p)\cdot f(p)$. This proves that the sum of two never-infinite measurable functions is measurable. A similar

proof is possible for the product of two measurable functions, although this case can be considered more readily by a combination of the previous theorem with one proving that the square of a measurable function is measurable. It may be thought for a moment that if $E(s)$, $F(s)$ are measurable for each s and if $E(s)$ is an increasing set-function, then

$$\int FdE$$

is measurable without further restrictions on F, but this is not true. For example, let $E(s)$ be the set of real numbers less than s, so that $\delta E(s)$ is the number (point) s itself. Let $f(t)$ be some non-measurable function of t. Denote by $F(s)$ the set of real numbers less than $f(s) + s - a$. Then $E(s)$, $F(s)$ are certainly measurable for every s. $F(s)\delta E(s)$ will be that number, if it exists, which is simultaneously equal to s and less than $f(s) + s - a$. Hence

$$\int F(s)dE(s)$$

is the set of numbers such that $s < f(s) + s - a$, that is to say, the set of numbers s for which $f(s) > a$. But $f(t)$ is non-measurable and therefore for some value of a the above set is non-measurable.

It would be interesting to study more closely the conditions which must be laid on $F(s)$ in order that the integral should be measurable. It is probably unnecessary that $F(s)$ should be continuous in one direction even in a number of intervals.

7. **Geometrical illustration.** It is helpful in a study of this integral to have in mind an illustration which is as follows: Let $E(s)$ be a set of values of the real variable x for each s. In the plane use Cartesian coördinates Ox horizontal, Os vertical where Oy is usually drawn. Through each point on Os draw a horizontal line and mark on it the points whose x-coördinates belong to $E(s)$. Then corresponding to the *set-function* $E(s)$ there is a plane set e. Similarly to $F(s)$ corresponds a plane set f.

If $E(s)$ is an increasing set-function, e consists of the points belonging to a collection of vertical lines which are unbounded above. If d is the plane set corresponding to $\delta E(s)$, d will consist of the lower bounds (where they exist) of these vertical lines. The integral $\int FdE$ consists of the projection on the x-axis of the plane set fd common to f and d. If $e(x)$ is some function of x and if $E(s)$ is the set of values of x for which $e(x) < s$, then the corresponding plane set e is the set of points above but not including the "graph" of $s = e(x)$. The plane set d corresponding to $\delta E(s)$ is the set of points of which the "graph" consists.

If now $F(s)$ is the set of values of x for which $f(x) < s$, the plane set f' corresponding to $F(s - t)$ (considered as a set-function of t) is obtained by taking the image of the curve $t = f(x)$ (in the xOt plane) with respect

to the x-axis, moving it up a constant distance s, and then by choosing all the points in the plane below but not including this transformed curve. The set $G(s)$ corresponding to $g(x) = e(x) + f(x)$ is the set of values of x for which the curve $t = e(x)$ falls strictly below the curve $t = s - f(x)$, that is to say, for which $e(x) + f(x) < s$.

Rice Institute,
Houston, Texas.

DIRICHLET'S PROBLEM.

By George E. Raynor.*

1. The main object of the following paper is to give a solution of Dirichlet's problem valid for less restricted types of boundaries than those hitherto considered. On the whole, the argument follows the classical lines closely and involves a compromise between the Schwarz alternating process and the Poincaré "Méthode du balayage." A large part of the paper may, therefore, be regarded as a simplified expository development of certain well-known theorems on potential theory. Although the problem is treated in three dimensions only, the method is equally applicable to n.

The writer here wishes to acknowledge his indebtedness to Professor J. W. Alexander, who has assisted him with numerous suggestions throughout the preparation of the paper.

2. For the purposes of this paper, a *region* R will be a set of points in three-space such that (1) to each point of the set there corresponds a sphere which encloses no point not of the set, (2) there exists a sphere enclosing all the points of the set, (3) given any two points P_1 and P_2 of the set, there is always a continuous arc $P(t)$, $t_1 \leqq t \leqq t_2$, made up of points of the set and joining P_1 to P_2: $P_1 = P(t_1)$, $P_2 = P(t_2)$. The *boundary* B of the region R will be the set of all limit points of the region which are not themselves points of the region. The set $R + B$ consisting of the points and boundary points of a region will thus be a closed set. A function F is said to be *continuous* on a set of points C if it has a finite value at every point of C and if to every point P of C and every number $\epsilon > 0$ there exists a number $\delta_{\epsilon P} > 0$ such that if P' be any point of C within a distance $\delta_{\epsilon P}$ of P,

$$|F(P) - F(P')| < \epsilon.$$

A function $V(x, y, z)$ is said to be *harmonic* in a region if at every point of the region it possesses first and second derivatives and if its second derivatives satisfy Laplace's equation

$$\Delta^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0.$$

* Presented to the American Mathematical Society, December 28, 1922.

DIRICHLET'S PROBLEM.

Let $U(x, y, z)$ be a function defined on the boundary B of a region R and continuous on B. The problem will be, if possible, to find a function $V(x, y, z)$ which is continuous over the domain $R + B$, harmonic in R, and identical with $U(x, y, z)$ on B.

During the course of the discussion, we shall also have occasion to deal with the following slight extension of this problem, though only in the case where the boundary B of the region R consists of a finite number of analytic surface elements. Let $U(x, y, z)$ be a function bounded on B and continuous at all points of B except along a finite number of analytic arcs A where $U(x, y, z)$ need not be defined. The problem will then be to find a function $V(x, y, z)$ which is bounded and continuous over the domain $R + B - A$, harmonic in R and identical with $U(x, y, z)$ in $B - A$.

3. In this section we shall prove a number of fundamental theorems concerning harmonic functions. Dirichlet's problem may be solved for the region interior to a sphere by means of Poisson's integral*

$$V(a, b, c) = \frac{1}{4\pi}\int\!\!\int_s U\frac{R^2 - \rho^2}{Rr^3}d\sigma \tag{1}$$

which defines the value of the required function V at any point *within* the sphere. In this formula the integral is extended over the surface of the sphere. R is the radius of the sphere, ρ the distance from the center to the point (a, b, c), r the distance from (a, b, c) to a variable point on the surface of the sphere and U is a continuous function of position on the surface of the sphere. The above integral is harmonic in a, b, c and is such that as (a, b, c) approaches a point of the surface in any manner whatever, $V(a, b, c)$ will approach the value of U at that point. The same integral also solves the extended Dirichlet problem when U is a bounded function continuous except perhaps along a finite number of analytic arcs. In fact, provided merely that the function U be bounded and integrable in the sense of Lebesgue, the integral (1) will define a function V harmonic within the sphere and such that as an interior point P approaches a point P_0 of the sphere at which U is continuous, the value $V(P)$ will approach the value $U(P_0)$.

If in formula (1) we put $\rho = 0$, r will become constant and equal to R and we shall obtain Gauss' mean value theorem

$$U(a_0, b_0, c_0) = \frac{1}{4\pi R^2}\int\!\!\int_s U d\sigma \tag{2}$$

* For a derivation of this formula see, for example, Goursat, Cours d'Analyse Mathématique, vol. 3, Chap. 28.

which gives the value of a harmonic function at the center (a_0, b_0, c_0) of a sphere as the average of its values on the surface. This formula shows at once that, if we consider the value of a function at an interior point of a region in which the function is harmonic, the values of the function in every small neighborhood of this point cannot be all greater or all less than the value of the function at that point. Hence we have at once the following theorem.

THEOREM 1. *If a function is harmonic in a region R, it can have neither a maximum nor a minimum in R.*

Here, we are, of course, using the terms maximum and minimum in the restricted sense.

THEOREM 2. *If a function V is harmonic in a region R with boundary B and continuous in the domain $R + B$, the greatest and least values of V in $R + B$ are attained on the boundary B.*

For a function which is continuous on a closed set of points is bounded and actually attains its least upper and greatest lower bounds.

THEOREM 3. *If the function V, harmonic in R and continuous in $R + B$, is constant (positive, negative) on B, it is constant (positive, negative) in $R + B$.*

THEOREM 4. *If V_1 and V_2 be functions harmonic in R and continuous in $R + B$ and if $V_1 = V_2$ ($V_1 > V_2$, $V_1 < V_2$) at every point of B, then $V_1 = V_2$ ($V_1 > V_2$, $V_1 < V_2$) at every point of $R + B$.*

This is seen on putting $V = V_1 - V_2$ in the previous theorem. In other words, we have

THEOREM 5. *If a solution of Dirichlet's problem exists, the solution is unique.*

We can also prove without difficulty that the extended Dirichlet problem referred to at the end of § 2 never admits of more than one solution. In other words,

THEOREM 6. *If V_1 and V_2 be two functions which are bounded and continuous in the domain $R + B - A$, harmonic in R and equal in $B - A$, the functions are identical in $R + B - A$.*

To prove the theorem we have only to show that the function $V = V_1 - V_2$ which is bounded and continuous in $R + B - A$, harmonic in R and zero on $B - A$ must vanish at all points of R. This we do with the aid of a comparison function. Let P_0 be any point of R and $V(P_0)$ the value of V at this point. We may assume without loss of generality that

$$V(P_0) \geqq 0,$$

for if $V(P_0)$ were negative we could work equally well with the function $-V$ instead of V. Moreover, since V is bounded, there exists a constant

M such that

$$V(P) < M$$

at all points of R. Now let μ be a positive constant and r the distance from a point of the system of arcs A to an arbitrary point (x, y, z) of space. Then the integral

$$I(P) = \mu \int \frac{ds}{r} \tag{5}$$

evaluated over all the arcs A defines the potential field due to a line distribution of density μ over A. Thus, $I(x, y, z)$ is positive and harmonic at every point P not of A and approaches infinity as the point P approaches a point of A.

Suppose now that the value of μ be chosen so small that at the point P_0,

$$I(P_0) < M.$$

Then the points P of R such that

$$I(P) < M$$

will form one or more sub-regions of R, one of which R' will contain the point P_0. Moreover, each point of the boundary of R' will either be a point of $B - A$ or a point of the equipotential surface

$$I(P) = M$$

of the function $I(P)$. Now at a boundary point of the first sort $I(P) > 0$ and $V(P) = 0$, while at a point of the second sort $I(P) = M$, $V(P) \leqq M$. Thus on the entire boundary of R' we shall have

$$I(P) \geqq V(P).$$

Therefore, by Theorem 4, since the functions $I(P)$ and $V(P)$ are both continuous on the boundary of R', this last relation is valid at every point within R' and in particular at the point P_0. It follows at once that the value of $V(P_0)$ cannot be positive; otherwise, by choosing the constant μ sufficiently small we could make

$$I(P_0) < V(P_0)$$

and thus be led to a contradiction.

As an immediate corollary to Theorem 6, we have the following theorem which will be needed later on in the discussion.

THEOREM 7. *If a function V be bounded and continuous in $R + B - A$, harmonic in R and non-negative on $B - A$, it is non-negative in $R + B - A$.*

This may be seen at once with the aid of the comparison function (5), where μ is now taken as a negative quantity which is allowed to approach zero.

4. From Poisson's integral we can derive a well-known inequality* which will be useful in proving the next theorem. In the formula

$$V(a, b, c) = \frac{1}{4\pi}\int\!\!\int_S U\frac{R^2 - \rho^2}{Rr^3}d\sigma$$

let $U \geqq 0$ everywhere on S. Then, by Theorem 3, V will be positive everywhere within S. Let V_0 be the value of V at the center of S and V_p its value at the point P. The maximum value of r is evidently $R + \rho$ and its minimum value $R - \rho$. We have then, replacing r by $R + \rho$,

$$V_p > \frac{1}{4\pi}\int\!\!\int_S U\frac{R^2 - \rho^2}{R(R + \rho)^3}d\sigma = \frac{1}{4\pi}\frac{R^2 - \rho^2}{R(R + \rho)^3}\int\!\!\int_S Ud\sigma.$$

But by Gauss' formula (2), we have

$$\int\!\!\int_S Ud\sigma = 4\pi R^2 V_0,$$

and hence finally

$$V_p > \frac{R(R - \rho)}{(R + \rho)^2}V_0. \tag{3}$$

In a similar manner, replacing r by $R - \rho$, we obtain

$$V_p < \frac{R(R + \rho)}{(R - \rho)^2}V_0. \tag{4}$$

Harnack's Theorem. *If a sequence of monotonic increasing functions, $u_1(x, y, z), \cdots, u_n(x, y, z), \cdots$, all of which are harmonic in a region R, converges at one point P of the region, it will converge at all points of R and the limit function will be harmonic in R.*

Let S be a sphere with center P and radius R lying entirely within the given region. Let A be any point in S at a distance ρ from P. If we consider the difference $(u_{n+p} - u_n)_P$, we have from the inequality (4) for all values of the indices n and p

$$(u_{n+p} - u_n)_A < \frac{R(R + \rho)}{(R - \rho)^2}(u_{n+p} - u_n)_P,$$

which proves the convergence of the sequence at the point A. The above expression also shows that within and on any sphere S' with the same center P but with radius $R' < R$ our sequence $u_1(x, y, z), \cdots, u_n(x, y, z), \cdots$ converges uniformly to a limit function u which may be written as a

* Cf., for example, Goursat, loc. cit.

uniformly convergent series,

$$\begin{aligned} u &= u_1 + (u_2 - u_1) + \cdots + (u_n - u_{n-1}) + \cdots \\ &= \sum (u_n - u_{n-1}), \qquad u_0 = 0. \end{aligned}$$

Replacing each term in this series by its value given by Poisson's integral for the sphere S' we have

$$\begin{aligned} u = \sum (u_n - u_{n-1}) &= \frac{1}{4\pi} \sum \int\int_{S'} (u_n' - u_{n-1}') \frac{R^2 - \rho^2}{Rr^3} d\sigma' \\ &= \frac{1}{4\pi} \int\int_{S'} \sum (u_n' - u_{n-1}') \frac{R^2 - \rho^2}{Rr^3} d\sigma', \end{aligned}$$

the signs $\sum$ and $\int\int$ being interchangeable since $\sum (u_n - u_{n-1})$ is *uniformly* convergent on S'. By our previous discussion we know that the last integral above is harmonic and we have that the limit function u is harmonic within S'.

Having established the theorem for the points in S' we shall now prove that it is true at any other point Q of the region R. Suppose that the theorem fails for the point Q. Join P and Q by a continuous arc $A(t)$, $t_1 \leqq t \leqq t_2$, $A(t_1) = P$, $A(t_2) = Q$, lying in the region R. Proceeding from P to Q along this arc we can then find a point R which is either the first point at which the theorem fails or is the last point which is such that the theorem is true for all points preceding it. But either of these situations is impossible, for if we take a sphere S'' lying entirely in the region R, having its center on the arc PR, and enclosing the point R, we have immediately by the first part of the proof that the theorem is true in this sphere and hence true for points immediately following R. Hence our supposition that the theorem fails at Q is false and the theorem is proved.

THEOREM 8. *If the sequence of functions $u_1(x, y, z), \cdots, u_n(x, y, z), \cdots$ defined in $R + B$ and harmonic in R converges uniformly everywhere on the boundary B of R, it will converge uniformly everywhere in $R + B$ and the limit function will be harmonic in R.*

Let $U_1, U_2, \cdots, U_n, \cdots$ be the values which $u_1, u_2, \cdots, u_n, \cdots$ take on the boundary B. Then by hypothesis if an $\epsilon > 0$ be given, we can find an m such that for $n \geqq m$ and for all positive values of p we will have at all points of B

$$|U_n - U_{n+p}| < \epsilon.$$

In particular this inequality holds for the maximum value of the left-hand member and hence we have by Theorem 2, at all interior points of R,

$$|u_n - u_{n+p}| < \epsilon,$$

which proves the uniform convergence.

That the limit function is harmonic in R can now be proved precisely as in the latter part of the previous theorem.

SCHWARZ'S ALTERNATING PROCESS.

5. This is a method whereby it is shown that if Dirichlet's problem has a solution for each of two overlapping regions R and R', then, under suitable conditions, it has a solution for the entire region $R + R'$ covered by the original pair of regions. It will be sufficient for our purposes to confine our attention to the case when the region R is the sum of the interiors of a finite number of spheres $S_1, S_2, \cdots, S_n$, no two of which are tangent to one another, and when the region R' consists of the interior of a single sphere S' such that S' is not tangent to any of the spheres $S_1, \cdots, S_n$. We shall also assume that the regions R and R' overlap but that neither contains the other. It is then to be proved that if the *extended* Dirichlet problem (§ 2) is always solvable for R it is always solvable for $R + R'$. We know, of course (§ 3), that the extended problem is solvable for R'.

Let B be the boundary of the region R and C the set of curves in which the boundary B intersects the boundary S' of R'. Moreover, let E and I be the portions of B exterior and interior to S' respectively and E' and I' the parts of S' exterior and interior to B respectively. Under certain conditions it may happen that either E or E' contains no points at all, that is to say, that the boundary of one of the two regions lies wholly within the other region. This will not invalidate the argument, however.

On the boundary of the region $R + R'$, a set of values $W(P)$ is given such that $W(P)$ is bounded,

$$|W(P)| < M,$$

and continuous over all of the boundary B with the possible exception of certain analytic arcs A. The problem is then to determine a solution of the extended Dirichlet problem for the region $R + R'$ corresponding to the arbitrary boundary value $W(P)$. Evidently we may assume without loss of generality that the function $W(P)$ is positive everywhere on $B - A$. For, as the function W is bounded, there exists a constant C such that $W + C$ is positive on $B - A$. Moreover, if we can solve the problem corresponding to the positive boundary values $W + C$, the required solution for the boundary values W will be obtained by merely subtracting the constant C from the previous solution.

Now let u_1 be a function harmonic in R, taking the assigned value W on E and the value zero on I. Then by Theorem 3 if the boundary values are continuous, or by Theorem 7 if they are discontinuous, the function u_1 will take a system of positive values on I'. Now let u_1' be a solution for

the region R' taking the values of u_1 on I' and the required values W on E'. The function u_1' will have a certain set of positive values on I, and we can form a new harmonic function for R taking the required values W on E and the values of u_1' on I. Proceeding in this manner by alternating back and forth from region R to region R' we obtain two sequences of functions,

$$u_1, u_2, \cdots, u_n, \cdots,$$
$$u_1', u_2', \cdots, u_n', \cdots,$$

the first set being harmonic and positive in R and taking the required values on E, and the second harmonic in R' and taking the proper values on E'.

Now we see from the manner in which these functions are obtained that at any point P in R the functions u are continually increasing. Furthermore, they are all bounded and hence approach a limit at P. Thus, by Harnack's Theorem, the functions $u_1, u_2, \cdots$ converge to a limit function u which is harmonic at all points of R. By the same argument we see that the sequence of u''s converges to a harmonic function u' in R'. In the region or regions bounded by I, I' and C the limits of the u and u' sequences must coincide with the limit of the monotonic increasing sequence

$$u_1, u_1', u_2, u_2', \cdots$$

and hence in this region we have $u = u'$. Thus, we may regard the limit function in R' as a continuation of the one in R and we have thus obtained a single function V harmonic in the region $R + R'$.

It now remains to be shown that the limit function $V(P)$ approaches the value $W(P)$ as the point P of $R + R'$ approaches a point P_0 of the boundary, provided P_0 is not on one of the arcs A. We first consider the case where the boundary point P_0 is a point of E. Since the boundary of the region $R + R'$ is made up of portions of spheres, no two of which are tangent to one another, a sufficiently small sphere S_0 about the point P_0 will certainly pass through points that are not of the region $R + R'$, as well as through points of the region itself. Moreover, if the sphere S_0 is made to shrink to the point P_0 by allowing its radius to approach zero, the ratio ρ between the area of the part of S_0 interior to $R + R'$ and the total area of S_0 will remain, from a certain point on, less than some constant q less than unity. If two of the spherical portions on the boundary of $R + R'$ were allowed to be tangent at P_0, the ratio in question would approach unity instead of remaining less than q, but the case of tangency we have explicitly ruled out.

We now construct a comparison function U_0 defined in the following

manner. Let S_0 be a sphere with center at P_0 and radius so small that the ratio ρ_0 of the portion of S_0 interior to $R + R'$ to the total area is less than q. Moreover, let N denote the least upper bound of the assigned boundary values W at points of $B - A$ within the sphere S_0 and $N + N'$ ($N' \geqq 0$) the least upper bound of the values of W at the points of $B - A$ as a whole. The comparison function U_0 is then to be such that at points of S_0 within $R + R'$ it takes on the value $N + N'$, at the remaining points of S_0 it takes on the value N, at points within S_0 it is harmonic and defined by means of a Poisson integral, using the boundary values just assigned on this surface S_0 itself. Thus at points P of S_0 interior to $R + R'$, we have

$$U_0(P) = N + N' \geqq u_n(P),$$

since no value of the function u_n can exceed the least upper bound of the assigned boundary values on $B - A$. Moreover, at points of $B - A$ interior to S_0,

$$U_0(P) \geqq N \geqq u_n(P) \qquad n = 1, 2, \cdots.$$

Therefore, by Theorem 7, the inequality

$$U_0(P) \geqq u_n(P) \qquad n = 1, 2, \cdots \tag{6}$$

holds at all points of the region or regions composed of the points of $R + R'$ interior to S_0. Consequently, a similar inequality holds for the limit function $V(P)$ also,

$$U_0(P) \geqq V(P) \tag{7}$$

at all points of $R + R'$ interior to S_0.

Now by Gauss' mean value theorem, the value of $U_0(P)$ at the center P_0 of S_0 is given by

$$U_0(P_0) = \rho_0(N + N') + (1 - \rho_0)N = N + \rho_0 N' < N + qN'. \tag{8}$$

Moreover, since the function U_0 is continuous within S_0, it will be possible to find a sphere S_1 interior to and concentric with S_0 such that within and on S_1 the inequality

$$U_0(P) < N + qN'$$

continues to hold. We are thus in a position to construct a second approximating function U_1, harmonic within S_1 and such that at points of S_1 interior to $R + R'$ the function U_1 takes on the value $N + qN'$, while at the remaining points of S_1, U_1 will take on the value N.

At points of S_1 interior to $R + R'$, we shall have for this new function

$$U_1(P) \geqq U_0(P) \geqq u_n(P),$$

while at points of B within or on S_1,

$$U_1(P) \geqq N \geqq u_n(P).$$

Consequently, for points of $R + R'$ interior or on S_1 we shall have the result

$$(6') \qquad U_1(P) \geqq u_n(P)$$

and therefore, also,

$$(7') \qquad U_1(P) \geqq V(P),$$

similar to (6) and (7) respectively. Moreover, the value of $U_1(P)$ at the center P_0 of S_1 is

$$U_1(P) = \rho_1(N + qN') + (1 - \rho_1)N = N + \rho_1 qN' < N + q^2N',$$

which is similar to (8). Consequently, there exists a third sphere S_2 interior to and concentric with S_1 within which we have

$$U_1(P) < N + q^2N'.$$

By repetitions of this argument, it is possible to find a sequence of spheres $S_0, S_1, S_2, S_3, \cdots$ about P_0 and a corresponding sequence of functions $U_0, U_1, U_2, U_3, \cdots$ such that within the sphere S_i the function U_{i-1} satisfies the inequality

$$U_{i-1}(P) < N + q^iN',$$

while at points of $R + R'$ within this sphere

$$U_i(P) > V(P).$$

Now, given any positive value ϵ, the initial sphere S_0 may be chosen so small that

$$N < W(P_0) + \frac{\epsilon}{2} \cdot$$

Moreover, the integer i may be chosen so large that

$$q^iN' < \frac{\epsilon}{2} \cdot$$

Thus

$$U_{i-1}(P) < W(P_0) + \epsilon$$

within S_i, and

$$(9) \qquad V(P) < W(P_0) + \epsilon$$

at all points of $R + R'$ within S_i.

In precisely the same way, using greatest lower bounds where before we used least upper bounds, we can prove the existence of a sphere S_i'

about P_0 within which

$$V(P) > W(P_0) - \epsilon. \tag{10}$$

Relations (9) and (10) establish the continuity of $V(P)$ at P_0. In similar fashion we can establish the continuity of the function $V(P)$ at a point P_0 of the portion E' of the boundary such that P_0 is not on an arc A. It only remains, therefore, to prove the continuity of $V(P)$ at a point P_0 of C which is not a point of A. This we do by constructing a series of spheres $S_0, S_1, \cdots$ about the point P_0 and the corresponding comparison functions $U_0, U_1, \cdots$, just as before. The only difference is that in treating this case, we have already established the continuity of the function $V(P)$ at all boundary points of $R + R'$ except those of A and C. Therefore, at each step we obtain the relation

$$U_i(P) > V(P)$$

directly from Theorem (7) without having to consider the approximating functions $u_n(P)$ or $v_n(P)$ at all. Thus, the extended Dirichlet problem for the region $R + R'$ is solved.

6. We are now in a position to establish Dirichlet's problem for a very general type of region R. It will be sufficient to assume that the boundary B of the region R is such that if a sphere S_0 of variable radius be drawn with center at any point P_0 of B, then the ratio ρ of the Lebesgue surface measure of the portion of the sphere interior to and on B to the total area of the sphere remains less than some constant $q(P_0) < 1$ as soon as the radius of the sphere is less than some value $r(P_0)$. This condition throws out of consideration a region R bounded by a surface B possessing an inward pointing spur of too sharp a type, though an inward pointing conical point is perfectly legitimate, or an outward pointing spur of any degree of sharpness. As a matter of fact, it is easy to prove that given a sufficiently sharp inward pointing spur on the boundary, the problem admits of no solution continuous at the tip of the spur. In the course of the discussion we shall see that the radius of S_0 need not shrink to zero continuously, it being sufficient merely that we can find for each point of B at least one denumerably infinite set of spheres satisfying the above conditions.

Before proceeding further we shall prove the following well-known lemma.*

LEMMA. *A three-dimensional region R can be covered by the interiors of a denumerably infinite set of spheres.*

*See, for example, Poincaré, "Sur les Équations aux Dérivées Partielles de la Physique Mathématique," in the Am. Jour. of Math., vol. 12, p. 211.

For, given any $\epsilon > 0$, the set σ_1 of all points of R within a distance of ϵ or more from the boundary B of R forms a closed set (which may be the null-set). Moreover, each point of σ_1 is the center of a sphere S of radius $\epsilon/2$ and such, therefore, that S neither meets nor contains a boundary point of R. Consequently, by the Heine-Borel theorem, the set σ_1 may be covered by a finite number of the spheres S. Consider next the infinite decreasing sequence of positive numbers

$$\epsilon, \frac{\epsilon}{2}, \frac{\epsilon}{3}, \cdots, \frac{\epsilon}{n}, \cdots.$$

The set of all points of R at a distance of at least ϵ/n but not more than $\epsilon/(n-1)$ from the nearest boundary point of R forms a closed set σ_n such that each point of σ_n is the center of a sphere of radius $\epsilon/(n+1)$ lying wholly in R. Hence by the same argument as before, the set σ_n may be covered by a finite number of spheres of the type required. We have thus constructed a denumerable set of sets σ_n each covered by a finite number of spheres and such that between them the sets σ_n include all the points of R. It therefore follows that the points of R may be covered by a denumerable number of finite sets of spheres, that is to say, by a denumerable number of spheres.

Now, let the covering spheres be arranged in a sequence as is always possible since they form a denumerable set. Each member of the sequence will then be preceded by a finite number of other spheres. By examining the spheres in the order 1, 2, 3, $\cdots$ and expanding each one slightly, though not enough for it to meet the boundary, we may always arrange so that no sphere is tangent to any of its predecessors. We shall assume in the sequel that the spheres have this property.

Now let $W(x, y, z)$ be a function defined and continuous everywhere on the boundary B of the region R. We shall assume that there exists a function $F(x, y, z)$ continuous in the domain $R + B$ and identical with $W(x, y, z)$ on B.* Let us now cover the region R with the interiors of a denumerably infinite set of spheres

$$S_1, S_2, \cdots, S_n, \cdots.$$

By means of Poisson's integral we can construct a function $v_1'(x, y, z)$ harmonic within S_1 and which takes on S_1 the same values as $F(x, y, z)$. We then define our first approximating function $v_1(x, y, z)$ to be equal to v_1' within and on S_1 and identical with $F(x, y, z)$ in the remaining portion of $R + B$. Now, if the spheres S_1 and S_2 intersect, by means of the

* For a proof of the existence of such a function, see an article by L. E. J. Brouwer in the Math. Annalen ,vol. 7, p. 209; or by Tietze, Jour. f. Math., vol. 145, p. 10.

alternating process we can find a function v_2' harmonic in the region R' covered by the interiors of S_1 and S_2 and taking on the boundary of this region the same values as $F(x, y, z)$. We then define a second approximating function v_2 as equal to v_2' in and on the boundary of R' and identical with $F(x, y, z)$ in the remaining portion of $R + B$. If S_1 and S_2 do not intersect, by Poisson's integral we can obtain two functions v_2' and v_2'' harmonic in S_1 and S_2 respectively, v_2' taking the same values as $F(x, y, z)$ on S_1 and v_2'' the same values as $F(x, y, z)$ on S_2. Then the function v_2 will be taken as equal to v_2' and v_2'' in and on S_1 and S_2 respectively and identical with $F(x, y, z)$ in the remaining portion of $R + B$. Proceeding in this manner, step by step, we obtain a sequence of functions,

$$v_1, v_2, \cdots, v_n, \cdots,$$

v_n being harmonic in the regions covered by the interiors of the first n spheres, taking on the boundaries of those regions the same values as $F(x, y, z)$ and identical with $F(x, y, z)$ in the remaining portion of $R + B$. We shall now prove that as n increases the function v_n approaches a limit function $v(x, y, z)$ which will be continuous in $R + B$, harmonic in R and identical with $W(x, y, z)$ on B, or, in other words, that the solution of Dirichlet's problem exists for the domain $R + B$.

Consider then a point P_0 of B and let W_0 be the value of $W(x, y, z)$ at this point. Let S_0 be a sphere with center P_0 and radius so small that the ratio ρ of the Lebesgue surface measure of the portion of S_0 within or on B to its total area is less than some constant $q < 1$. Let B_n' be the boundary of the region R' within which the approximating function v_n is harmonic and M be the least upper bound of $F(x, y, z)$ in $R + B$. Within B_n' we know, by Theorem 2, that v_n is less than the greatest value of $F(x, y, z)$ on B_n'. Hence, since v_n is identical with $F(x, y, z)$ in the portion of $R + B$ on and exterior to B_n', we have

$$v_n(x, y, z) \leqq M \tag{11}$$

at all points of $R + B$. Let M' be the least upper bound of $F(x, y, z)$ within or on S_0. Let us now construct, by means of Poisson's integral, a comparison function U_0 harmonic within S_0 and taking on the portion of S_0 interior to B the value M and on the portion exterior to B the value M'. We then have at once that within S_0

$$U_0 \geqq M' \geqq F(x, y, z). \tag{12}$$

Now, if B_n' intersects S_0, the portion of R' within S_0 will be made up of regions bounded partly by S_0 and partly by B_n' on which $v_n = F(x, y, z)$.

Hence on the boundaries of these regions we have

$$U_0 \geqq v_n, \tag{13}$$

and by Theorem 4 the same relation will subsist within these regions. In the portions of the region $R - R'$ within S_0 we have

$$v_n = F(x, y, z), \tag{14}$$

and hence by (12) we find (13) holding for this region. On the other hand, if B_n' does not intersect S_0 we once more obtain (13) directly from (14) and (12). Hence in each case we see that the comparison function U_0 is greater than all the approximating functions within S_0.

We may now take a sequence of spheres S_n with center P_0 and with radii decreasing to zero and set up, precisely as described in connection with the alternating process, a sequence of comparison functions

$$U_0, U_1, \cdots, U_n, \cdots$$

such that in S_n, U_n will be greater than all of the approximating functions. Since $F(x, y, z)$ is continuous in $R + B$, given an $\epsilon > 0$ the initial sphere S_0 may be taken so small that within this sphere M' will differ from W_0 by less than $\epsilon/2$. Then, since our decreasing spheres are subject to exactly the same condition as in the preceding section, we can take n so large that ultimately in S_n, U_n will differ from the value W_0 by less than ϵ. Hence all the approximating functions will be less than $W_0 + \epsilon$ in S_n. Now by using greatest lower bounds where before we used least upper bounds we obtain by an exactly analogous argument that all the approximating functions will be greater than $W_0 - \epsilon$ in some sphere S_n' with center at P_0. Hence if we let S_{nP_0}'' be a sphere with center at P_0 and interior to both S_n and S_n', we have the result, given any $\epsilon > 0$ we can find for each point P of B a sphere S_{nP}'' with P as center within which the oscillations of the approximating functions v_n all remain less than 2ϵ. From this set of spheres we can choose by the Heine-Borel Theorem a finite sub-set which will cover the boundary B. Consider now any point P' in the region R and draw about it a small sphere S' lying entirely in R. Let us now choose a value m of n so large that the boundary B_n' of the region in which the approximating function v_n is harmonic lies entirely in the above sub-set of spheres and encloses the sphere S'. By the above argument we have that for all values of $n \geqq m$ the oscillations of the approximating functions will be less than ϵ everywhere on B_n' and hence by Theorem 2 will be less than ϵ on S'. Therefore, by Theorem 8 the approximating functions converge to a limit in S' which is harmonic in S'. Hence in particular, we have that at any interior point P' of the region R the

approximating functions converge to a limit, and this limit function is harmonic at P'.

It now remains to prove that our limit function v takes on the assigned boundary values $W(x, y, z)$ on B. But this follows at once by precisely the same argument as was used in the case of the alternating process to show that the limit function there obtained approaches the proper boundary values. We have precisely the same inequalities subsisting between the comparison functions and the limit functions, and by the restriction on the boundary B made at the beginning of this section, we have the same condition on the decreasing set of spheres for each point of B. Hence we have finally the result,

THEOREM 9. *Dirichlet's problem has a solution for every region R whose boundary B is such that, if a sphere be drawn about any point of B, the ratio of the measure of the points of the surface of the sphere interior to and on B to the whole area of the sphere will ultimately remain less than unity as the radius of the sphere approaches zero. The shrinking process need not be continuous but may be made by a denumerably infinite set of steps only.*

ANNIHILATORS OF MODULAR INVARIANTS AND COVARIANTS.*

By Olive C. Hazlett.

Introduction.

1. **Abstract and relation to the literature.** Up to the present, not very much has been accomplished toward developing a theory of modular covariants for the general case—i.e., for the general form or for the general field. Dickson has proved that modular invariants and covariants of any system of binary forms possess the finiteness property when the coefficients of the transformations are marks of any Galois Field $GF[p^n]$ of order p^n,† and fundamental sets of invariants, seminvariants and covariants have been found by various writers for the more important special cases.‡ But in these latter papers, one is struck by the fact that the methods used are ones which apply admirably to the cases considered but in some way fail of complete generality.§

Nevertheless, the results for the different special cases have analogies, some of which are rather striking. Some of these analogies are shown in the conditions that a given function φ be a seminvariant of a form f and in the closely related subject of annihilators of invariants and covariants. A few years ago, Professor Dickson ‖ found annihilators of modular semin-

* Read before the American Mathematical Society, September 6, 1920. The work of this paper has been facilitated by the purchase of an abstract journal and other books from a grant made by the American Association for the Advancement of Science.

Since finishing this MS., there has appeared an article by W. L. G. Williams on "Formal modular seminvariants" (presented to the American Mathematical Society October 30, 1920; published in the Transactions of the American Mathematical Society for January, 1921), in which he proves Theorem III of the present paper. Nevertheless, I have decided to leave my own paper unchanged, especially since his proof does not seem to me very convincing (see § 6).

† "General theory of modular invariants," Trans. Amer. Math. Soc., vol. 10 (1909), pp. 123–158; "Proof of the finiteness of modular covariants," ibid., vol. 14 (1913), pp. 299–310.

‡ Dickson's results are summarized in his Madison Colloquium Lectures, "On invariants and the theory of numbers" (1914). Glenn, "A fundamental system of formal covariants modulo 2 of the binary cubic," Trans. Amer. Math. Soc., vol. 19 (1918), pp. 109–118; "Modular concomitant scales with a fundamental system of formal covariants, modulo 3, of the binary quadratic," ibid., vol. 20 (1919), pp. 154–168.

§ Perhaps this statement should be qualified. The methods of constructing invariants and covariants have generality in that they are applicable to other cases and some of them are applicable even for general modulus p or for the general form f; but none of them is applicable to all covariants of every form.

‖ "Invariants of binary forms under modular transformations," Trans. Amer. Math. Soc., vol. 8 (1907), pp. 205–232.

variants of the binary quadratic and binary cubic analogous (in a general way) to those in classical invariant theory. Then he found a set of annihilators for modular seminvariants of a binary quadratic (or cubic) for some of the Galois Fields $GF[p^n]$ of order p^n, where p is a small prime and n is greater than 1. This gives at once necessary conditions that a polynomial φ be a seminvariant of f. For the cases considered, he verified that these conditions are also sufficient.*

The present paper attacks the problem in a slightly different way and obtains results which apply to any system of binary forms and any Galois Field $GF[p^n]$ of order p^n. The annihilators of modular invariants obtained in this manner are of the type anticipated in the paper by Professor Dickson; so also are the set of necessary and sufficient conditions that a polynomial φ be a modular seminvariant. It is interesting to note that these operators are also annihilators of formal modular invariants if no reductions are made by Galois' generalization of Fermat's Theorem ($a^{p^n} \equiv a$). Hence, since the modular covariants of a system S may be obtained from the modular invariants of an enlarged system S', we readily have annihilators of modular covariants. In the same manner, we obtain annihilators of formal modular covariants. These annihilators lead to a set of n necessary and sufficient conditions that a polynomial φ be a modular covariant (formal or otherwise).

2. **Summary of previous results.** The first published work on annihilators of modular invariants was in a paper by Dickson. As he there pointed out, the differential operators which annihilate an invariant are more complicated in the theory of modular invariants than in the theory of classic invariants, for in a series of powers of an arbitrary mark t of the $GF[p^n]$, certain terms now combine—namely, t^i, $t^{i+\mu}$, $t^{i+2\mu}$, $\cdots$, where $\mu = p^n - 1$. Bearing this fact in mind, and applying Taylor's Theorem, he finds annihilators for some special cases.

For the first example, he considers the form

$$a_0x^2 + a_1xy + a_2y^2, \tag{1}$$

in which the coefficients are integers reduced modulo 3. Let φ be a polynomial in the a's with all exponents ≤ 2. There is no loss of generality in doing this, since any polynomial may be reduced to this form by applying Fermat's Theorem, which in this case gives us $a^3 \equiv a$ (mod 3). Under the transformation

$$\begin{aligned} x &= x' + ty', \\ y &= y', \end{aligned} \tag{2}$$

let a_1 and a_2 be transformed into $a_1 + \alpha_1$ and $a_2 + \alpha_2$ respectively, and

* For an outline of his results, the reader is referred to § 2 of the present paper.

let φ be transformed into φ'. Then, by Taylor's Theorem, $\varphi' - \varphi$ is readily expressed as a polynomial in α_1 and α_2, in which the exponents of the α's are $\leqq 2$ and the coefficients are partial derivatives of φ with respect to a_1 and a_2 divided by an integer relatively prime to 3.

By rearranging terms, it is evident that

$$\varphi' - \varphi = t\delta_1\varphi + t^2\delta_2\varphi,$$

where $\delta_1\varphi$ and $\delta_2\varphi$ are differential operators on φ. A necessary condition that $\varphi' \equiv \varphi$ is clearly $\delta_1\varphi \equiv 0$. It is not so evident, however, that this is also sufficient. In the classical case, $\delta_2\varphi$ is readily shown to be $\frac{1}{2}\delta_1(\delta_1\varphi) = \frac{1}{2}\delta_1{}^2\varphi$. The classic procedure does not obtain here because δ_1 is a differential operator which applies only to a polynomial in which the exponents are all $\leqq 2$, whereas the polynomial $\delta_1\varphi$ does not have all its exponents $\leqq 2$. Hence we have no right to talk about $\delta_1(\delta_1\varphi)$ in this case. If we let $[\delta_1\varphi]$ denote $\delta_1\varphi$ in reduced form, i.e., with every exponent $\leqq 2$, then $\delta_1[\delta_1\varphi] \equiv \frac{1}{2}\delta_2\varphi$. Hence it follows that $\delta_1\varphi \equiv 0$ is both a necessary and sufficient condition that $\varphi' \equiv \varphi$; that is, φ is a seminvariant of the binary quadratic modulo 3 if and only if $\delta_1\varphi \equiv 0$.

Dickson found that a similar statement can be made about the seminvariants of the other special cases that he studied, provided that the field is a Galois Field $GF[p^n]$ of order p where p is a prime.

But, just as soon as we consider seminvariants of a quantic in the $GF[p^n]$, where $n > 1$, then difficulties arise. If ψ is a polynomial in t of degree $\leqq \mu = p^n - 1$, then

$$\psi(a + t) - \psi(a) = t\psi'(a) + \frac{t^2}{2!}\psi''(a) + \cdots + \frac{1}{i!}t^i\psi^{(i)}(a) + \cdots + \cdots + \frac{1}{\mu!}t^\mu\psi^{(\mu)}(a).$$

When $n > 1$, the denominators $i!$ are not all relatively prime to p. Consider the quantic

$$a_0x^m + a_1x^{m-1}y + a_2x^{m-2}y^2 + \cdots + a_my^m \tag{3}$$

and subject x and y to the transformation

$$\begin{aligned} x &= x' + ty', \\ y &= y'. \end{aligned} \qquad (t \text{ in the } GF[p^n]). \tag{4}$$

If φ is any polynomial in the a's, and φ' denotes its transform, then

$$\varphi' - \varphi \equiv t\delta_1\varphi + t^2\delta_2\varphi + \cdots + t^i\delta_i\varphi + \cdots + t^\mu\delta_\mu\varphi, \tag{5}$$

where the δ_i are differential operators. Professor Dickson considered special quantics when $n > 1$, and from his results he conjectured that in general necessary and sufficient conditions that φ shall be an invariant

of (3) under the group of transformations (4) are the vanishing of $\delta_i\varphi$ where $i = 1, p, p^2, \cdots, p^{n-1}$.

Nine years later, Professor Glenn* observed that, for the $GF[p]$, any formal covariant would have to be annihilated by

$$(6)\qquad \Upsilon = O_0 + O_1 x \frac{\partial}{\partial y} + O_2 x^2 \frac{\partial^2}{\partial y^2} + \cdots + O_j x^j \frac{\partial^j}{\partial y^j} + \cdots.$$

He does not say what the O_j are except that they are partial differential operators "in the derivatives with respect to the coefficients of (3), non-homogeneous as to the derivatives the orders of which range from zero to infinity in each O_j." In this paper he points out that, if we proceed as in the proof of Robert's Theorem on the unique-determination of an algebraic covariant from its leader, we get relations among the coefficients of the covariant and the operators O_j which are not recurrent as they are in the classic case. Under special conditions, he gives these relations; but they are so complicated that they do not seem particularly useful.

Annihilators for $GF[p]$.

3. **A preliminary formula.** Consider a system S of forms with coefficients, the a's, which are marks of the Galois Field $GF[p^n]$ of order p^n. Let φ be any polynomial of the a's. When we subject the variables x and y to the transformation

$$(7)\qquad \begin{aligned} x &= x' + ty' \\ y &= y' \end{aligned} \qquad (t \text{ any mark in } GF[p^n]),$$

let the a's be transformed into the a''s and let φ be transformed into φ'. Then, by the Lie theory,

$$(8)\qquad \varphi' - \varphi = t\,\Omega\varphi + \frac{t^2}{2!}\Omega^2\varphi + \cdots + \frac{t^k}{k!}\Omega^k\varphi + \cdots,$$

where

$$\Omega\varphi = \frac{\partial\varphi'}{\partial t}\bigg|_{t=0}, \qquad \Omega^2\varphi = \Omega(\Omega\varphi) = \frac{\partial^2\varphi'}{\partial t^2}\bigg|_{t=0},$$

and, in general,

$$\Omega^k\varphi = \Omega(\Omega^{k-1}\varphi) = \frac{\partial^k\varphi'}{\partial t^k}\bigg|_{t=0}.$$

Note that Ω is the Aronhold annihilator for the case when the a's and t are in the field of ordinary complex numbers. Hence, for the classic case, a necessary and sufficient condition that $\varphi' = \varphi$ is that $\Omega\varphi = 0$. This is not true, however, when t and the a's are marks of a finite field, since then

* "The formal modular invariant theory of binary quantics," Trans. Amer. Math. Soc., vol. 17 (1916), pp. 545–556, especially pp. 547–548.

the different powers of t are not distinct and thus certain terms coalesce. If t is any mark of the Galois Field $GF[p^n]$ of order p^n, then $t^{p^n} \equiv t$ in the field (by Galois' generalization of Fermat's Theorem). Thus, in the field, we have the congruence

$$(9) \qquad \varphi' - \varphi \equiv t \sum_{k=0}^{\infty} \frac{\Omega^{1+k(p^n-1)}}{[1 + k(p^n - 1)]!} + t^2 \sum_{k=0}^{\infty} \frac{\Omega^{2+k(p^n-1)}}{[2 + k(p^n - 1)]!} + \cdots + t^{p^n-1} \sum_{k=0}^{\infty} \frac{\Omega^{(k+1)(p^n-1)}}{[(k + 1)(p^n - 1)]!}.$$

In each of these coefficients, it must be borne in mind that the division by the factorial is purely a formal one. This is legitimate even when the factorial is not relatively prime to p, since $\partial^q \varphi / \partial t^q$ is always exactly divisible by $q!$. We shall write (9) for convenience in the form

$$(9') \qquad \varphi' - \varphi = t\delta_1\varphi + t^2\delta_2\varphi + \cdots + t^\mu\delta_\mu\varphi,$$

where $\mu = p^n - 1$.

4. **Significance of the differential operators.** Before proceeding any further, it will be well to pause a moment to consider the full force of the operators δ_k. When the a's and t are independent variables, then $\varphi' - \varphi$ is given by (8), where Ω is the classical Aronhold annihilator denoted by U in Lie's theory. But when the a's and t are indeterminate marks of the $GF[p^n]$, then φ' may be written in a variety of ways and for each such way of writing φ' we get a different expansion for $\varphi' - \varphi$. When φ' denotes the result of replacing the a's by the a''s in a purely formal manner—i.e., without reducing by the aid of Fermat's Theorem, we shall say φ' is written in unreduced form. If we go to the extreme of reducing the exponents of the t so that they are all $\leq \mu = p^n - 1$, then

$$(10) \qquad \varphi' - \varphi \equiv t \left.\frac{\partial \varphi'}{\partial t}\right|_{t=0} + \frac{t^2}{2!} \left.\frac{\partial^2 \varphi'}{\partial t^2}\right|_{t=0} + \cdots + \frac{t^\mu}{\mu!} \left.\frac{\partial^\mu \varphi'}{\partial t^\mu}\right|_{t=0}.$$

But $\left.\dfrac{\partial \varphi'}{\partial t}\right|_{t=0}$ is not now the same as $\Omega\varphi$ and the derivatives of higher order are not now the same as the corresponding iterations of Ω. In fact, in the coefficients of t we will now have not only the coefficient of t when φ' is written in unreduced form, but also the coefficient of t^{1+p^n-1}, of $t^{1+2(p^n-1)}$, etc. Hence the coefficient of t in (10) is the sum of a number of expressions—viz., the partial derivatives of the unreduced φ' with respect to t of orders 1, $1 + (p^n - 1)$, $1 + 2(p^n - 1)$, $\cdots$. A similar statement applies to the coefficients of the other powers of t in (10).

But $\varphi' - \varphi$ is also given by the congruence (9). Hence, in (9), the coefficient of t is actually congruent to the first partial derivative of φ'

with respect to t if we reduce by Fermat's Theorem before differentiating; the coefficient of t^2 is congruent to $1/2!$ times the second partial derivative of φ' with respect to t if we reduce by Fermat's Theorem first; and so on. Since this is true of any polynomial φ' in the a's, then we would expect that the coefficient of t^2 would be found by applying the operator

$$D_1 = \frac{\partial}{\partial t} + \frac{1}{[1 + (p^n - 1)]!}\frac{\partial^{1+(p^n-1)}}{\partial t} + \cdots = \sum_{k=0}^{\infty} \frac{1}{[1 + k\mu]!}\frac{\partial^{1+k\mu}}{\partial t^{1+k\mu}}$$

to $D_1\varphi'$ and then setting $t = 0$ in the result. This will be proved in the sequel.

5. **Annihilators of modular invariants for** $GF[3]$. First we shall consider the case when the field consists of the classes of residues of integers taken modulo 3. Then (9′) becomes

$$\varphi' - \varphi \equiv t\delta_1\varphi + t^2\delta_2\varphi,$$

where

$$\delta_1 = \Omega + \frac{1}{3!}\Omega^3 + \frac{1}{5!}\Omega^5 + \cdots + \frac{1}{(1 + 2k)!}\Omega^{1+2k} + \cdots$$

and

$$\delta_2 = \frac{1}{2!}\Omega^2 + \frac{1}{4!}\Omega^4 + \frac{1}{6!}\Omega^6 + \cdots + \frac{1}{[2(k + 1)]!}\Omega^{2(k+1)} + \cdots.$$

As indicated in § 4, we are led to suspect that $\delta_2 = \delta_1^2/2!$; at least our hope is enough to warrant our computing δ_1^2. Now

$$\begin{aligned}\delta_1^2 &= \left[\sum_{k=0}^{\infty}\frac{1}{(1 + 2k)!}\Omega^{1+2k}\right]\left[\sum_{l=0}^{\infty}\frac{1}{(1 + 2l)!}\Omega^{1+2l}\right]\\
&= \frac{1}{1!}\Omega^2 + \left[\frac{1}{1!\,3!} + \frac{1}{3!\,1!}\right]\Omega^4 + \left[\frac{1}{1!\,5!} + \frac{1}{3!\,3!} + \frac{1}{3!\,1!}\right]\Omega^6 + \cdots\\
&\qquad + \left[\frac{1}{1!\,(1 + i2)!} + \frac{1}{3!\,[1 + (i - 1)2]!}\right.\\
&\qquad\qquad \left. + \frac{1}{5!\,[1 + (i - 2)2]!} + \cdots\right]\Omega^{2(i+1)} + \cdots.\end{aligned}$$

It is easy to verify for small values of i that the expression inside the braces in each coefficient is identically congruent modulo 3 to $2!$ times the corresponding coefficient in the expression for δ_2 and then proceed by induction. We leave the details to the reader. Thus we have

$$\varphi' - \varphi \equiv t\delta_1\varphi + \frac{t^2}{2!}\delta_1^2\varphi,$$

and it at once follows that a necessary and sufficient condition that $\varphi' \equiv \varphi$ (mod 3) is that $\delta_1\varphi \equiv 0$ (mod 3).

By interchanging x and y and also interchanging x' and y', it follows that a polynomial φ in the a's is unaltered under the group of transformations

$$(11)\qquad \begin{aligned} x &= x' \\ y &= tx' + y' \end{aligned} \qquad (t, \text{ any mark in } GF[3]),$$

if and only if

$$(12)\qquad \delta_1' = O + \frac{1}{3!}O^3 + \frac{1}{5!}O^5 + \cdots + \frac{1}{(1+2k)!}O^{1+2k} + \cdots,$$

in which O is the second Aronhold annihilator of the classic theory.

6. **Annihilators of modular invariants for** $GF[p]$. For the set of classes of residues of integers taken modulo a general prime p,*

$$(13)\qquad \varphi' - \varphi = t\delta_1\varphi + t^2\delta_2\varphi + \cdots + t^{p-1}\delta_{p-1}\varphi,$$

where

$$\delta_k = \sum_{l=0}^{\infty} \frac{1}{[k + l(p-1)]!}\Omega^{k+l(p-1)} \qquad (k = 1, \cdots, p-1).$$

Here

$$\delta_1^2 = \left[\sum_{l=0}^{\infty} \frac{1}{[1 + l(p-1)]!}\Omega^{1+l(p-1)}\right]\left[\sum_{r=0}^{\infty} \frac{1}{[1 + r(p-1)]!}\Omega^{1+r(p-1)}\right]$$
$$= \sum_{s=0}^{\infty}\left[\sum_{q=0}^{s} {}_{2+s(p-1)}C_{1+q(p-1)}\right]\frac{1}{[2 + s(p-1)]!}\Omega^{2+s(p-1)}.$$

Now the expression inside the brackets is congruent to 2 modulo p. For, since by Fermat's Theorem

$$(x + y)^{2+s(p-1)} \equiv (x + y)^2 = x^2 + 2xy + y^2 \qquad (\text{mod } p),$$

whenever x and y are integers, the sum of the coefficients of all terms of the form $x^{1+q(p-1)}y^{1+r(p-1)}$ must be congruent to 2 modulo p. Thus $\delta_1^2 \equiv 2\delta_2 \pmod{p}$.

Also

$$\delta_1^3 \equiv \delta_1(\delta_1^2) \equiv 2\delta_1\delta_2$$
$$\equiv 2\left[\sum_{l=0}^{\infty} \frac{1}{[1 + l(p-1)]!}\Omega^{1+l(p-1)}\right] \times \left[\sum_{r=0}^{\infty} \frac{1}{[2 + r(p-1)]!}\Omega^{2+r(p-1)}\right]$$

* Dr. Williams reasons thus: "A necessary and sufficient condition that φ' be independent of t and so $\equiv \varphi$, modulo p, which it is when $t \equiv 0$, modulo p, is that $\partial\varphi'/\partial t \equiv \delta_1\varphi$, whence the theorem follows" (Transactions, vol. 22, p. 60). Every part of this statement is self-evident with the exception of the assertion that a sufficient condition that $\varphi' - \varphi \equiv 0 \pmod{p}$ is that $\partial\varphi'/\partial t \equiv 0 \pmod{p}$. Although such a statement is well known to be true when the field of definition is infinite and φ is a continuous function of t where t ranges over a continuous interval (a, b), I do not see how one is thereby justified in omitting the proof that the coefficients of the higher powers of t in $\varphi' - \varphi$ are actually the iterations of $\delta_1\varphi$ multiplied by suitable constants, even when the field is finite. It is, however, very easy to give a careful proof.

$$\equiv 2\sum_{s=0}^{\infty}\left[\sum_{q=0}^{s} {}_{3+s(p-1)}C_{1+q(p-1)}\right]\frac{1}{[3+s(p-1)]!}\Omega^{3+s(p-1)}.$$

In this last member, the expression inside the brackets is congruent to 3 (mod p), since it is the coefficient of x^2y in $(x+y)^{3+s(p-1)}$ when all exponents are reduced modulo $p-1$. Thus

$$\delta_1^3 \equiv 2\cdot 3\sum_{s=0}^{\infty}\frac{1}{[3+s(p-1)]!}\Omega^{3+s(p-1)} \equiv 3!\,\delta_3.$$

In general, by induction, we have

$$\begin{aligned}\delta_1^k &\equiv \delta_1(\delta_1^{k-1})\\ &\equiv \left[\sum_{l=0}^{\infty}\frac{1}{[1+l(p-1)]!}\Omega^{1+l(p-1)}\right]\\ &\qquad\times\left[(k-1)!\sum_{r=0}^{\infty}\frac{1}{[(k-1)+r(p-1)]!}\Omega^{(k-1)+r(p-1)}\right]\\ &\equiv (k-1)!\sum_{s=0}^{\infty}\left[\sum_{q=0}^{s} {}_{k+s(p-1)}C_{1+q(p-1)}\right]\\ &\qquad\qquad\times\frac{1}{[k+s(p-1)]!}\Omega^{k+s(p-1)}.\end{aligned}\tag{15}$$

But the expression inside the last pair of brackets is the sum of the coefficients of all terms of the form

$$x^{(k-1)+(s-q)(p-1)}y^{1+q(p-1)}\tag{16}$$

in the expansion of $(x+y)^{k+s(p-1)}$. But, when x and y are any two integers,

$$(x+y)^{k+s(p-1)} \equiv (x+y)^k = x^k + kx^{k-1}y + \cdots \pmod{p};$$

and thus, since all terms of the form (16) in $(x+y)^{k+s(p-1)}$ coalesce to give the term $x^{k-1}y$ in $(x+y)^k$,

$$\sum_{q=0}^{s} {}_{k+s(p-1)}C_{1+q(p-1)} \equiv k.$$

Therefore (15) gives

$$\delta_1^k \equiv k!\sum_{s=0}^{\infty}\frac{1}{[k+s(p-1)]!}\Omega^{k+s(p-1)} = k!\,\delta_k.\tag{17}$$

Since (17) holds when k has any value from 1 to $p-1$ inclusive, (13) becomes

$$\varphi' - \varphi \equiv t\delta_1\varphi + \frac{t^2}{2!}\delta_1^2\varphi + \cdots + \frac{t^{p-1}}{(p-1)!}\delta_1^{p-1}\varphi.\tag{18}$$

Hence a necessary and sufficient condition that $\varphi' \equiv \varphi$ is that $\delta_1\varphi \equiv 0$ in the field. Thus we have proved

THEOREM I. *Let S be a system of forms in the variables x and y with coefficients, the a's, which may assume any set of values which are integers, reduced modulo p, a prime. Let φ be a polynomial in the a's. Then a necessary and sufficient condition that φ be a modular invariant under the group of transformations $x = x' + ty'$, $y = y'$ is that $\delta_1\varphi \equiv 0$ (mod p). Here δ_1 is the differential operator*

$$\Omega + \frac{1}{p!}\Omega^p + \frac{1}{[1 + 2(p-1)]!}\Omega^{1+2(p-1)} + \cdots + \frac{1}{[1 + k(p-1)]!}\Omega^{1+k(p-1)} + \cdots,$$

in which Ω is the (first) Aronhold annihilator used in the classic invariant theory. It must be remembered that, since the a's are integers reduced modulo p, this theorem requires merely that $\delta_1\varphi$ shall vanish after all possible reductions have been made via Fermat's Theorem.

7. **A second annihilator for modular invariants in $GF[p]$.** By interchanging the rôles of x and y in the above theorem, we have

THEOREM II. *Let S and φ be defined as in Theorem I. Then a necessary and sufficient condition that φ be a modular invariant under the group of transformations $x = x'$, $y = tx' + y'$ is that $\delta_1'\varphi \equiv 0$ (mod p) where*

$$\delta_1'\varphi \equiv O + \frac{1}{p!}O^p + \frac{1}{[1 + 2(p-1)]!}O^{1+2(p-1)} + \cdots + \frac{1}{[1 + k(p-1)]!}O^{1+k(p-1)} + \cdots.$$

Here O is the second Aronhold annihilator used in the theory of classic invariants and is given by the formula

$$O\varphi = \sum_i \frac{\partial\varphi}{\partial a_i}\left(\frac{\partial a_i'}{\partial t}\right)_{t=0}.$$

8. **Two annihilators for formal invariants in $GF[p]$.** In the last three sections, we have been considering modular invariants of S that were not (necessarily) formal and thus the a's were integers reduced modulo p. Accordingly, the condition that φ be unaltered under the transformation $x = x' + ty'$, $y = y'$ was that $\delta_1\varphi$ shall vanish whenever the a's are in the field. If we now turn our attention to the corresponding problem for formal modular invariants, the a's are no longer integers reduced modulo p, but are independent variables and consequently a^{p^n}

is no longer congruent to a. If we follow through the reasoning of sections 3 to 7, it is evident that the results still hold, provided all the work is purely formal and we do not replace a^p by a. Thus we have

THEOREM III. *Let S be a system of forms in the variables x and y with coefficients, the a's, which are independent variables; and let φ be a polynomial in the a's. Then a necessary and sufficient condition that φ be a formal modular invariant of the system S under the group of transformations $x = x' + ty'$, $y = y'$, where t is any integer reduced modulo p, is that $\delta_1\varphi$ be identically congruent to zero, modulo p.*

In the same manner, we readily prove the

COROLLARY. *If the coefficients of some forms of the system S, say the a's, are independent variables while the coefficients of the other forms of S are integers reduced modulo p, then φ is a modular invariant of S under the group of transformations $x = x' + ty'$, $y = y'$ (where t is any integer taken modulo p) which is formally invariant under the group as to the $\bar{a}$'s if and only if $\delta_1\varphi \equiv 0$, where this congruence is an identity in the $\bar{a}$'s.*

By interchanging the rôles of x and y in Theorem III, we have

THEOREM IV. *Let S and φ be defined as in Theorem III. Then a necessary and sufficient condition that φ be a formal modular invariant under the group of transformations $x = x'$, $y = tx' + y'$ (where t is any integer reduced modulo p) is that $\delta_1'\varphi$ be identically congruent to zero, modulo p.*

9. **Two annihilators of modular covariants for $GF[p]$.** By the aid of Theorems III and IV we can readily derive two annihilators of modular covariants (whether formal or otherwise). For it has already been shown that every modular covariant of the system $S(x, y)$—with variables x and y—can be obtained in a simple manner from the modular invariants of the system S' consisting of the forms of $S(\xi, \eta)$ and the additional linear $\eta x - \xi y$—in which the variables are ξ and η.* For every modular covariant of $S(x, y)$ is a polynomial in $L = x^p y - xy^p$ and in the modular invariants M of S' which have been made formally invariant as to x and y.

By the Corollary of Theorem III, a function φ is a modular invariant of $S'(\xi, \eta)$ under the group of transformations induced by $x = x' + ty'$, $y = y'$ and is formally invariant as to x and y under the group if and only if it is annihilated by

$$\Delta_1 = \Omega' + \frac{1}{p!}\Omega'^p + \frac{1}{[1 + 2(p-1)]!}\Omega'^{1+2(p-1)} + \cdots$$

$$= \sum_{k=0}^{\infty} \frac{1}{[1 + k(p-1)]!}\Omega'^{1+k(p-1)},$$

where $\Omega' = \Omega - y\dfrac{\partial}{\partial x}$. Since L is itself a modular invariant of $S'(\xi, \eta)$—

* Trans. Amer. Math. Soc., vol. 21 (1920), p. 253.

namely, the invariant which is zero for all classes of $S'(\xi, \eta)$—which has been made formally invariant as to x and y, this last statement applies also to L or to any polynomial in L and in the modular invariants M which have been made formally invariant as to x and y. Moreover, a necessary and sufficient condition that φ be a formal modular covariant of S is that $\Delta_1\varphi$ be identically congruent to zero. Thus we have

THEOREM V. *Let S be a system of forms in the variables x and y with coefficients, the a's, which may assume any set of values which are integers, reduced modulo p, a prime. Let φ be a polynomial in the a's and in x and y. Then a necessary and sufficient condition that φ be a modular covariant under the group of transformations $x = x' + ty'$, $y = y'$ (where t is any integer reduced modulo p) is that $\Delta_1\varphi \equiv 0 \pmod{p}$. This congruence must be an identity in the variables x and y. If, in addition, it is an identity in the a's, then φ is a formal modular covariant of S.*

If, in Theorem V, we interchange the rôles of x and y, we have

THEOREM VI. *Let S and φ be defined as in Theorem V. Then a necessary and sufficient condition that φ be a modular covariant of S under the group of transformations $x = x'$, $y = tx' + y'$ is that $\Delta_1'\varphi \equiv 0 \pmod{p}$. This congruence must be an identity in x and y. When the coefficients of S are independent variables, then φ is a formal covariant of S under this group of transformations if and only if $\Delta_1'\varphi \equiv 0 \pmod{p}$, this congruence being an identity in the a's and in the variables x and y.*

Generalization to $GF[p^n]$.

10. **Annihilators for invariants.** Thus far, t—the coefficient of the transformation—has been an integer reduced modulo some prime p; now we shall generalize and consider the case when t is a polynomial in some variable (say i) reduced modulis a polynomial $P(i)$ of degree n and some prime p. Thus t is congruent to one of the p^n expressions of the form $c_0i^{n-1} + c_1i^{n-2} + \cdots + c_{n-1}i + c_n$, where the c's range independently over the set of integers $0, 1, \cdots, p-1$. If $P(i)$ is irreducible modulo p, then the set of all such expressions is closed under the processes of addition, subtraction, multiplication and division (provided the divisor is not zero modulis $P(i)$ and p), and it is called the Galois Field of order p^n—denoted by $GF[p^n]$. For any mark a of this field, we have holding Galois' generalization of Fermat's Theorem, $a^{p^n} \equiv a \pmod{P(i), p}$.

Let S be, as before, a system of forms in the variables x and y with coefficients, the a's, which may be independent variables or indeterminate marks of the $GF[p^n]$. A polynomial φ in the a's is invariant under the group of transformations

$$x = x' + ty' \qquad (t, \text{ a mark of } GF[p^n])$$
$$y = y'$$

if and only if the increment of φ is congruent to zero in the field. But, by § 3,

$$\varphi' - \varphi = \sum_{k=1}^{\mu} t^k \delta_k \varphi, \qquad (\mu = p^n - 1),$$

where

$$\delta_k \varphi = \sum_{l=0}^{\infty} \frac{1}{[k + l(p^n - 1)]!} \Omega^{k+l(p^n-1)} \qquad (k = 1, \cdots, p^n - 1).$$

This is congruent to zero in the field if and only if each $\delta_k \varphi \equiv 0$.

Now in the work on annihilators for the $GF[p^n]$, we have already shown that

$$(19) \qquad \delta_1^k \equiv k!\, \delta_k \qquad (\text{mod } p)$$

when $k = 1, \cdots, p - 1$. These results hold here without change, so that, if $\delta_1 \varphi = 0$, then $\delta_k \varphi \equiv 0$ when $k < p$. But, when $k = p$, although we still have (19) holding, it does not now follow that, if $\delta_1 \varphi \equiv 0$, then so also is $\delta_p \varphi \equiv 0$. For, since $\delta_p \varphi$ is formally congruent to $\delta_1^p \varphi / p!$ in the field, then the vanishing of $\delta_p \varphi$ requires that $\delta_1^p \varphi \equiv 0$ [mod $pP(i)$ and p^2]. Thus, when $n > 1$, there arises a second necessary condition which is independent of the first.

Since the result of dividing by $p + l$—where $l = 1, \cdots, p - 1$—and then reducing modulo p is the same as the result obtained by dividing by l and then reducing modulo p, it follows that, when $0 \leq k < p$,

$$\delta_{k+p} = \frac{\delta_1^{k+p}}{(k+p)!} = \frac{\delta_1^k}{k!} \frac{(\delta_1^p)}{p!} \equiv \frac{\delta_1^k}{k!} (\delta_p) \qquad (\text{mod } p).$$

Hence, if $\delta_p \varphi \equiv 0$, then $\delta_{k+p} \varphi \equiv 0$ in the field for $0 \leq k < p$. When $k = p$, we have

$$\left[\frac{\delta_1^p}{p!}\right]^2 = \left[\sum_m \frac{1}{[p + m(p^n - 1)]!} \Omega^{p+m(p^n-1)}\right]^2$$
$$= \sum_{m=0}^{\infty} C_m^2 \frac{1}{[2p + m(p^n - 1)]!} \frac{\partial^{2p+m(p^n-1)}}{\partial t^{2p+m(p^n-1)}},$$

in which

$$C_m^2 = \sum_{k=0}^{m} {}_{2p+m(p^n-1)}C_{p+k(p^n-1)}.$$

Now

$$\sum_{k=0}^{m} {}_{2p+m(p^n-1)}C_{p+m(p^n-1)} \equiv {}_{2p}C_p \qquad (\text{mod } p)$$

since

$$(x + y)^{2p+m(p^n-1)} \equiv (x + y)^{2p},$$

whenever x and y are marks in the field $GF[p^n]$. But ${}_{2p}C_p \equiv {}_2C_1$ modulo p, since $\frac{2p-l}{p-l} \equiv 1$ (modulo p) when $l \not\equiv 0$ (modulo p). Hence

$$\frac{1}{2!}\left[\frac{\delta_1{}^p}{p!}\right]^2 \equiv \delta_{2p} \qquad (\text{mod } p).$$

By induction, we see that

$$(20) \qquad \left[\frac{\delta_1{}^p}{p!}\right]^l \equiv (l-1)!\sum_{m=0}^{\infty} C_m{}^l \frac{\partial^{lp+m(p^n-1)}}{\partial t^{lp+m(p^n-1)}} \qquad (\text{mod } p)$$

$$\equiv l!\, \delta_{lp}. \qquad (\text{mod } p).$$

Hence necessary and sufficient conditions that $\delta_k\varphi$ $(k = 1, 2, \cdots, p^2-1)$ shall vanish are that $\delta_1\varphi$ and $\delta_p\varphi$ shall vanish.

More generally, by induction, we see that if $s = k_0 + k_1p + \cdots + k_{n-1}p^{n-1}$ (each k an integer between 0 and $p-1$), then

$$\frac{\delta_1{}^s}{s!} = \frac{1}{(k_0)!}[\delta_1]^{k_0}\frac{1}{(k_1)!}\left[\frac{\delta_1{}^p}{p!}\right]^{k_1}\cdots\frac{1}{(k_{n-1})!}\left[\frac{\delta_1{}^{p^{n-1}}}{(p^{n-1})!}\right]^{k_{n-1}}.$$

Hence we readily prove

THEOREM VII. *Let S be a system of forms in the variables x and y with coefficients, the a's, which may be independent variables or may be indeterminates ranging over the Galois Field $GF[p^n]$ of order p^n. Let φ be a polynomial in the a's. Then necessary and sufficient conditions that φ be an invariant of S under the group of transformations $x = x' + ty'$, $y = y'$ (where t is any mark of $GF[p^n]$) are that φ be annihilated in the field by δ_k where $k = 1, p, p^2, \cdots, p^{n-1}$. Moreover φ is a formal invariant of S if and only if these congruences hold identically in the field when the a's are independent variables.*

In a similar manner, we find necessary and sufficient conditions that φ shall be an invariant of S under the group of transformations $x = x'$, $y = tx' + y'$ (where t is any mark of the field $GF[p^n]$).

11. **Annihilators for covariants.** We can now readily prove the analogous theorem for covariants, for the modular covariants of a system S are the modular invariants of an enlarged system S'. We leave the details of the proof to the reader since they are very similar to those given in § 9. Thus we have

THEOREM VIII. *Let S be a system of forms as in Theorem VII and let φ be a polynomial in the a's and in x and y. Then necessary and sufficient conditions that φ be a covariant of S under the group of transformations $x = x' + ty'$, $y = y'$ (where t is any mark of $GF[p^n]$) are that φ be annihilated in the field by Δ_k where $k = 1, p, p^2, \cdots, p^{n-1}$. Here Δ_1 is defined*

as in § 9, and $\Delta_k = \frac{1}{k!}(\Delta_1)^k$. *Moreover,* φ *is a formal covariant of* S *if and only if these congruences* $\Delta_k\varphi \equiv 0$ *hold identically in the field when the* a*'s are independent variables.*

In a similar manner, we derive necessary and sufficient conditions that φ shall be a covariant of S under the group of transformations $x = x'$, $y = tx' + y'$ (where t is any mark of the field $GF[p^n]$).

MOUNT HOLYOKE COLLEGE,
SOUTH HADLEY, MASS.

SYSTEMS OF LINEAR INEQUALITIES.

BY WALTER B. CARVER.

In a paper under this same title,* Professor L. L. Dines found a necessary and sufficient condition for the existence of solutions of a system of linear inequalities, for both the homogeneous and non-homogeneous cases. His condition was expressed in terms of the " I-rank " of the matrix. It is the purpose of the present paper to give, in a quite different form, a necessary and sufficient condition for the *non-existence* of solutions; and to consider the questions of the independence of a system and the equivalence of two systems.

Let S represent the system of m linear inequalities in n variables,

$$\sum_{j=1}^{n} \alpha_{ij} x_j + \beta_i > 0, \qquad i = 1, 2, \cdots m,$$

in which the β's may or may not all be zero. For brevity we may write

$$L_i(x) \quad \text{for} \quad \sum_{j=1}^{n} \alpha_{ij} x_j + \beta_i \quad \text{and} \quad L_i'(x) \quad \text{for} \quad \sum_{j=1}^{n} \alpha_{ij} x_j.$$

The matrix of the coefficients, $\| \alpha_{ij} \|$ (not including the β's), will be denoted by M.

A system of inequalities will be said to be *consistent* or *inconsistent* according as solutions of the system do or do not exist. A single inequality will be inconsistent only when

$$\alpha_{i1} = \alpha_{i2} = \cdots = \alpha_{in} = 0, \qquad \text{and} \qquad \beta_i \leqq 0.$$

THEOREM 1. *If for a system S the rank of the matrix M is m, the system is consistent.*

We may suppose that the non-vanishing determinant of order m in the matrix M is made up of the first m columns of the matrix; and consider the set of equations,

$$\sum_{j=1}^{m} \alpha_{ij} x_j = c_i, \qquad i = 1, 2, \cdots m.$$

Since the determinant of the coefficients does not vanish, solutions of this set of equations exist for any values of the c's. Fix c's satisfying the relations $c_i > -\beta_i$, and let $a_1, a_2, \cdots a_m$ be the solution of the resulting set of equations. Then evidently $a_1, a_2, \cdots a_m, 0, \cdots 0$ is a solution of the system S of inequalities.

* These Annals, vol. 20, p. 191.

A system S of m inequalities will be said to be *irreducibly inconsistent* when the system S is inconsistent, but each sub-system of $m - 1$ inequalities in S is consistent; i.e., when the omission of any one inequality from the inconsistent system leaves a consistent system. A single inequality will be irreducibly inconsistent if it is inconsistent.

THEOREM 2. *If the system S is irreducibly inconsistent, there exists a set of constants $k_1, k_2, \cdots k_{m+1}$, homogeneously unique, such that*

$$\sum_{i=1}^{m} k_i L_i(x) + k_{m+1} \equiv 0,$$

*$k_1, k_2, \cdots k_m$ being positive and k_{m+1} positive or zero; and the rank of the matrix M must be $m - 1$.**

By hypothesis there exists a set of numbers $a_1, a_2, \cdots a_n$ or, briefly, a point† a, which satisfies all the inequalities except the first one. This may be conveniently expressed by saying that there exists a point a which gives the row of m symbols

$$\bar{0} \quad + \quad + \quad + \quad \cdot \quad \cdot \quad \cdot \quad +;$$

the double symbol "$\bar{0}$" indicating that $L_1(a)$ is either negative or zero, and the following plus signs indicating that each of the expressions $L_i(a)$, for $i \neq 1$, is positive. Similarly, there exists a point giving each of the rows

$$\begin{array}{cccccccc} + & \bar{0} & + & + & \cdot & \cdot & \cdot & +, \\ + & + & \bar{0} & + & \cdot & \cdot & \cdot & +, \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\,, \\ + & + & + & + & \cdot & \cdot & \cdot & \bar{0}. \end{array}$$

If h_1 and h_2 are any two positive numbers whose sum is unity, we may speak of the point $h_1a + h_2b$ (i.e., the set of numbers $h_1a_1 + h_2b_1$, $h_1a_2 + h_2b_2, \cdots h_1a_n + h_2b_n$) as a point *between* a and b. Since the expressions $L_i(x)$ are linear, $L_i(h_1a + h_2b) = h_1L_i(a) + h_2L_i(b)$. Suppose now that a point b should exist which, when substituted in the L's, makes at least one of them positive and all of them either positive or zero; giving, for instance,

$$+ \quad 0 \quad 0 \quad + \quad \cdot \quad \cdot \quad \cdot \quad +.$$

Since there is a point a which gives

$$\bar{0} \quad + \quad + \quad + \quad \cdot \quad \cdot \quad \cdot \quad +,$$

* The method of proof of this theorem was suggested to the author by Professor Hurwitz.

† Whether the system S is or is not homogeneous, the set of numbers indicated by the phrase "the point a" will not be a homogeneous set; i.e., the point a does not mean the set of numbers $ca_1, ca_2, \cdots ca_n$.

it is evident that there would be a point between a and b which would make the L's all positive. But this is contrary to the hypothesis that the system S is inconsistent. Hence every point which, when substituted in the L's, makes at least one of them positive, will also make at least one of them negative. It follows that where we used the double symbol "$\bar{0}$" above, the zero can not occur; and that there are therefore points giving each of the rows

$$\begin{array}{cccccccc} - & + & + & + & \cdot & \cdot & \cdot & +, \\ + & - & + & + & \cdot & \cdot & \cdot & +, \\ + & + & - & + & \cdot & \cdot & \cdot & +, \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot, \\ + & + & + & + & \cdot & \cdot & \cdot & -. \end{array}$$

Again, if the points a and b give respectively

$$- \quad + \quad + \quad + \quad \cdot \quad \cdot \quad \cdot \quad +$$

and

$$+ \quad - \quad + \quad + \quad \cdot \quad \cdot \quad \cdot \quad +,$$

then some point between a and b will make $L_1(x)$ vanish, and will give

$$0 \quad - \quad + \quad + \quad \cdot \quad \cdot \quad \cdot \quad +.$$

This point must make $L_2(x)$ negative, as indicated, because we have shown that a point which makes any of the L's positive must make at least one L negative. Evidently, then, there exists a point which makes any arbitrarily chosen L vanish, any other one negative, and all the rest positive.

Between the two points which give respectively

$$0 \quad - \quad + \quad + \quad \cdot \quad \cdot \quad \cdot \quad +$$

and

$$0 \quad + \quad - \quad + \quad \cdot \quad \cdot \quad \cdot \quad +,$$

there is similarly some point which gives

$$0 \quad 0 \quad - \quad + \quad \cdot \quad \cdot \quad \cdot \quad +.$$

By continuing this process, it is evident that we can establish the existence of a point p such that

$$L_i(p) = 0, \qquad i \neq s, t; \qquad L_s(p) < 0, \qquad \text{and} \qquad L_t(p) > 0,$$

L_s and L_t being any two of the L's chosen arbitrarily. Also, by carrying

the process one step further, it may be shown that there exists a point q such that

$$L_i(q) = 0, \qquad i \neq s; \qquad \text{and} \qquad L_s(q) \gtreqless 0.$$

If, now, there exists a set of constants $k_1, k_2, \cdots k_{m+1}$, not all zero, such that

$$\sum_{i=1}^{m} k_i L_i(x) + k_{m+1} \equiv 0,$$

it is evident that the identity

$$\sum_{i=1}^{m} k_i L_i'(x) \equiv 0$$

must also hold, and that k_{m+1} must be equal to $-\sum_{i=1}^{i=m} k_i\beta_i$. Since the system S is inconsistent, the rank of the matrix M must be less than m (by theorem 1); and hence it follows that there is at least one set of constants $k_1, k_2, \cdots k_m$, not all zero, such that

$$\sum_{i=1}^{m} k_i L_i'(x) \equiv 0.$$

If we first suppose that our system S is homogeneous, $L_i'(x) \equiv L_i(x)$, and we have

$$\sum_{i=1}^{m} k_i L_i(x) \equiv 0.$$

Substituting the point p in this identity, we have

$$k_s L_s(p) + k_t L_t(p) = 0;$$

and since $L_s(p) < 0$ and $L_t(p) > 0$, it follows that either k_s and k_t are both zero, or neither of them is zero and they have the same sign. But these are *any* two constants of the set $k_1, k_2, \cdots k_m$; and since not all of them are zero, none of them are zero and they all have the same sign. They may evidently all be made positive, and $k_1, k_2, \cdots k_m, 0$ is then such a set of constants as our theorem requires.

Suppose, on the other hand, that the system S is non-homogeneous. Then $L_i'(x) \equiv L_i(x) - \beta_i$, and we have

$$\sum_{i=1}^{m} k_i L_i(x) \equiv \sum_{i=1}^{m} k_i\beta_i.$$

Substituting the point q in this identity, we have

$$k_s L_s(q) = \sum_{i=1}^{m} k_i\beta_i.$$

If $\sum k_i\beta_i \neq 0$, then $k_s \neq 0$ and differs in sign from $\sum k_i\beta_i$. This means that none of the k's are zero, and that all of them have the sign contrary

to that of $\sum k_i\beta_i$. We may make them all positive, and with $k_{m+1} = -\sum k_i\beta_i$ we have a set of constants of the kind required by the theorem. For the case $\sum k_i\beta_i = 0$, none of the k's are zero, by the same argument that was used in the homogeneous case. Hence there must be a point q such that $L_i(q) = 0$, $i = 1, 2, \cdots m$. It follows that the transformation

$$x_j = x_j' + q_j, \qquad j = 1, 2, \cdots n,$$

sends this non-homogeneous system into the corresponding homogeneous system. And since such a transformation does not affect the existence or non-existence of solutions, the corresponding homogeneous system must be irreducibly inconsistent. It follows then, from our treatment of the homogeneous case, that the set of constants $k_1, k_2, \cdots k_m$ all have the same sign. Taking them all positive, and putting $k_{m+1} = 0$, we have such a set as the theorem requires.

We have then shown that for any system, homogeneous or non-homogeneous, which is irreducibly inconsistent, there exists at least one set of constants, $k_1, k_2, \cdots k_m$, not all zero, such that

$$\sum_{i=1}^{m} k_i L_i'(x) \equiv 0;$$

that in any such set none of the constants is zero, and all of them may be taken as positive; and that when we adjoin to any such set

$$k_{m+1} = -\sum_{i=1}^{m} k_i\beta_i$$

we then have such a set of k's as our theorem requires. It will follow that this set of constants is homogeneously unique when we show that the rank of the matrix M must be $m - 1$.

Suppose that the rank r of the matrix M were less than $m - 1$. Then for a properly chosen sub-set of $r + 1$ of the inequalities, say the first $r + 1$ of them, there would be a set of constants $f_1, f_2, \cdots f_{r+1}$, not all zero, such that

$$\sum_{i=1}^{r+1} f_i L_i'(x) \equiv 0.$$

These $r + 1$'s, together with $m - r - 1$ zeros, would make up a set of k's such that

$$\sum_{i=1}^{m} k_i L'(x) \equiv 0.$$

But we have shown that one of such a set of k's can not vanish unless they all vanish. Hence the rank of the matrix M can not be less than, and must therefore be equal to, $m - 1$. And it follows that the set of k's is homogeneously unique. This completes the proof of the theorem.

It is rather obvious that if solutions exist for a homogeneous system, they exist for any corresponding non-homogeneous system; and that the converse is not true.* But it follows from the proof of the last theorem that if a non-homogeneous system is irreducibly inconsistent, the same will be true of the corresponding homogeneous system.

Another by-product of the proof of the last theorem is the following fact: If in the matrix M of an irreducibly inconsistent system S we pick out any non-vanishing determinant of order $m - 1$, and throw out all the columns of the matrix except those involved in this determinant, we have left a matrix of $m - 1$ columns and m rows, in which the m determinants of order $m - 1$ alternate in sign, none of them vanishing.

THEOREM 3. *A necessary and sufficient condition that a given system S be inconsistent is that there should exist a set of $m + 1$ constants, k_1, k_2, $\cdots k_{m+1}$, such that*

$$\sum_{i=1}^{m} k_i L_i(x) + k_{m+1} \equiv 0,$$

at least one of the k's being positive, and none of them being negative.

As to the sufficiency of the condition: suppose that a point a is a solution of the system, i.e., that $L_i(a) > 0$, $i = 1, 2, \cdots m$. Since at least one k is positive, and none are negative, it is obvious that the identity

$$\sum_{i=1}^{m} k_i L_i(x) + k_{m+1} \equiv 0$$

could not hold for this point. Hence there can be no solutions.

It remains to establish the necessity of the condition. If the system S is inconsistent, but not irreducibly inconsistent, we may drop out some inequality from the system which will leave an inconsistent sub-system of $m - 1$ inequalities. If this sub-system is not irreducibly inconsistent, we may drop one inequality from it, leaving an inconsistent sub-system of $m - 2$ inequalities. By continuing this process, we must finally arrive at an irreducibly inconsistent sub-system of ρ inequalities, where $1 \leqq \rho \leqq m$. We may think of this sub-system as consisting of the first ρ of the inequalities of our system S; and, by theorem 2, we have a set of constants $k_1, k_2, \cdots k_\rho, k_{m+1}$, such that

$$\sum_{i=1}^{\rho} k_i L_i(x) + k_{m+1} \equiv 0,$$

$k_1, k_2, \cdots k_\rho$ being positive, and k_{m+1} positive or zero. If now we put $k_{\rho+1} = k_{\rho+2} = \cdots = k_m = 0$, we have the set of constants required by our theorem.

In connection with the above proof it may be noted that an inconsistent

* Cf. Dines, loc. cit.

system S may have a number of different irreducibly inconsistent sub-systems. The rank of the matrix of any such sub-system of ρ inequalities is $\rho - 1$, and can not be greater than the rank r of the matrix M. Hence we must always have $\rho \leqq r + 1$. For a given inconsistent system, there may or may not be an irreducibly inconsistent sub-system containing as many as $r + 1$ inequalities.*

An inequality will be said to be *superfluous* in a system S, in which $m \geqq 2$, when it is satisfied by every point which satisfies all the other inequalities of the system.† In an inconsistent system, $m \geqq 2$, an inequality can be superfluous if and only if the sub-system obtained by omitting this inequality is inconsistent. We therefore have at once

THEOREM 4. *The necessary and sufficient condition that the inequality* $L_s(x) > 0$ *should be superfluous in an inconsistent system* S *is that there should exist a set of constants* $k_1, k_2, \cdots k_{m+1}$, *such that*

$$\sum_{i=1}^{m} k_i L_i(x) + k_{m+1} \equiv 0,$$

with $k_s = 0$, *at least one* k *positive, and none negative.*

THEOREM 5. *The necessary and sufficient condition that the inequality* $L_s(x) > 0$ *should be superfluous in a consistent system* S *is that there should exist a set of constants* $k_1, k_2, \cdots k_{m+1}$, *such that*

$$\sum_{i=1}^{m} k_i L_i(x) + k_{m+1} \equiv 0,$$

k_s *and no other* k *being negative, and at least one* k *being positive.*

The sufficiency of the condition is rather obvious. We have by hypothesis

$$L_s(x) \equiv \frac{k_1}{-k_s} L_1(x) + \cdots + \frac{k_{s-1}}{-k_s} L_{s-1}(x) + \frac{k_{s+1}}{-k_s} L_{s+1}(x) + \cdots + \frac{k_m}{-k_s} L_m(x) + \frac{k_{m+1}}{-k_s}.$$

---

* For instance, for the system

(1) $x_1 > 0$, (2) $x_2 > 0$, (3) $-2x_1 - x_2 - 5 > 0$, (4) $4x_1 + 2x_2 + 1 > 0$,

for which $r = 2$, if we drop (4), we have at once an irreducibly inconsistent system with $\rho = 3$; but if we first drop (1), we must then drop (2) before we arrive at an irreducibly inconsistent system with $\rho = 2$. Again, in the system

$x_1 > 0$, $x_2 - 1 > 0$, $x_3 - 2 > 0$, $-x_2 - x_3 + 1 > 0$,

for which $r = 3$, we can drop only the first inequality, giving $\rho = 3$.

† For the case $m = 1$, we shall define an inequality to be superfluous in the system consisting of itself alone when and only when it is an identical inequality, i.e., when all the coefficients of the variables are zero and the constant term is positive. It is readily verified that the necessary and sufficient conditions of the next two theorems are in accord with this definition.

where at least one coefficient on the right is positive and none are negative. If in this identity we substitute a point a which satisfies all the inequalities of the system except possibly $L_s(x) > 0$, we see at once that $L_s(a)$ must also be positive.

To prove the necessity of the condition, consider a system S' obtained by replacing the inequality $L_s(x) > 0$ in S by the contradictory inequality $-L_s(x) > 0$. By hypothesis, every point which satisfies the inequalities of S other than $L_s(x) > 0$ must also satisfy this inequality, and hence can not satisfy the inequality $-L_s(x) > 0$. Hence S' is inconsistent, and there exists a set of constants $k_1, k_2, \cdots k_{m+1}$, such that

$$k_1L_1(x) + \cdots + k_{s-1}L_{s-1}(x) + k_s\{-L_s(x)\} + \cdots k_mL_m(x) + k_{m+1} \equiv 0,$$

at least one k being positive, and none negative. Moreover, we know that $k_s \neq 0$, and that at least one other k does not vanish, for otherwise the system S would be inconsistent. If then we replace k_s by $-k_s$, we have the set of constants required by the theorem.

A system S will be said to be *independent* if it contains no superfluous inequalities. In accordance with this definition, an irreducibly inconsistent system is an inconsistent system which is independent. A single inequality will always be independent except in the case of the identical inequality noted above.

Two systems may be said to be *equivalent* if every point which satisfies either of them satisfies the other one. Any two inconsistent systems are equivalent, and an inconsistent system can not be equivalent to a consistent system. A single inequality is obviously equivalent to another single inequality when and only when they are identically the same except possibly for a positive constant factor.

THEOREM 6. *If two systems S and $\sum$, each of which is independent and consistent, are equivalent, the number of inequalities in the two systems is the same, and each inequality of one system is equivalent to one and only one inequality of the other system; i.e., the inequalities of the two systems are identical except for possible positive constant factors.*

Let $L_s(x) > 0$ be any inequality of the system S. Since it is not superfluous in S, and S is consistent, there must exist a point a such that $L_i(a) > 0$, $i \neq s$, and $L_s(a) \leqq 0$; and also a point b such that $L_i(b) > 0$, $i = 1, 2, \cdots m$. Hence there must be a point c coincident with a or between a and b, such that $L_i(c) > 0$, $i \neq s$, and $L_s(c) = 0$. Since there is one such point, there must be an infinite number of them; for every point satisfying the equation $L_s(x) = 0$ and lying in a sufficiently small region about c will satisfy the same conditions. Let G represent the set of all points satisfying these conditions, $L_i(c) > 0$, $i \neq s$, and $L_s(c) = 0$. Let H represent the set of all points satisfying the system S. The only

limit points of H which do not belong to H are points which satisfy the equations $L_i(x) = 0$ for one or more values of i, and the inequalities $L_i(x) > 0$ for the remaining values of i. The points of the set G are such limit points of H. But since, by hypothesis, H is also the set of all points satisfying the system $\sum$, each point of the set G must satisfy at least one equation $\lambda_i(x) = 0$ corresponding to an inequality $\lambda_i(x) > 0$ of the set $\sum$. And since there are only a finite number of inequalities in the set $\sum$, at least one equation, say $\lambda_s(x) = 0$, must be satisfied by an infinite number of points of G. Hence the equation $\lambda_s(x) = 0$ must be equivalent to the equation $L_s(x) = 0$; and the inequality $\lambda_s(x) > 0$ must be equivalent to the inequality $L_s(x) > 0$. Moreover, an inequality of S can not be equivalent to more than one inequality of $\sum$, for in that case these inequalities in $\sum$ would all be equivalent to each other, and all but one of them would be superfluous in $\sum$.

If one drops a superfluous inequality from a consistent system S, the remaining system of $m - 1$ inequalities is evidently equivalent to the original system. If this system of $m - 1$ inequalities is not independent, a superfluous inequality may be dropped from it. By continuing this process, we must finally arrive at an independent sub-system equivalent to the original system.* The order in which the superfluous inequalities are dropped in this process is immaterial; for, by the last theorem, any two independent sub-systems obtained in this way can differ only by positive constant factors in the inequalities. This is in distinct contrast to the facts for an inconsistent system.

ITHACA, N. Y.

* The only exception is the trivial case in which all the inequalities of the system S are the identical inequalities noted above.

EULER SQUARES.

By Harris F. MacNeish.

1. **Introduction.** Euler Squares were first considered in a paper, "Recherches sur une espèce de carrés magique," Commentationes Arithmeticæ Collectæ, 1849, vol. II, pp. 302–361. In this paper Euler proposed the following problem now well known as "The problem of the 36 officers."* Six officers of six different ranks are chosen from each of six different regiments. It is required to arrange them in a solid square so that no officer of the same rank or of the same regiment shall be in the same row or in the same column. The problem is equivalent to that of arranging 36 pairs of integers, each less than or equal to six, in a square array so that the first (or second) numbers of the pairs in any row or column are all distinct, and no two pairs are identical. Such a square array would be called an Euler Square of index 6, 2.

In this paper we shall be concerned with more general squares defined as follows. An Euler square of order n, degree k and index n, k is a square array of n^2 k-ads of numbers, $(a_{ij1}, a_{ij2}, \cdots, a_{ijk})$, where $a_{ijr} = 1, 2, \cdots, n$; $r = 1, 2, \cdots, k$; $i, j = 1, 2, \cdots, n$; $n > k$; $a_{ipr} \neq a_{iqr}$ and $a_{pjr} \neq a_{qjr}$ for $p \neq q$ and $a_{ijr}a_{ijs} \neq a_{pqr}a_{pqs}$ for $i \neq p$ and $j \neq q$.

The impossibility of constructing squares of index n, 2 for $n \equiv 2 \pmod 4$ was stated without proof by Euler in the paper referred to above. A very laborious proof for index 6, 2 obtained by combining two squares of index 6, 1 has been given by G. Tarry (Mathesis, vol. 20, July, 1901). A geometrical proof by methods of Analysis Situs has been given by J. Petersen (Annuaire des Mathématiciens, 1901–02, pp. 413–426). A third method is given for index n, 2, $n \equiv 2 \pmod 4$, by P. Wernicke, "Das Problem der 36 Offiziere," Jahresbericht der deutschen Mathematiker-Vereinigung, vol. 19, 1910, p. 264. The method of Wernicke is proved to be incorrect in an article under the same title in the same journal, vol. 31, 1922, p. 151, by H. F. MacNeish. An Euler Square of degree one is called a Latin Square and of degree two a Græco-Latin Square.

We shall show how to construct Euler squares for the following cases: (A) Index p, $p - 1$ for p prime; (B) Index p^n, $p^n - 1$ for p prime; (C) Index n, k, where $n = 2^r p_1^{r_1} p_2^{r_2} \cdots$ for $p_1, p_2, \cdots$ distinct odd primes and where $k + 1$ equals the least of the numbers 2^r, $p_1^{r_1}$, $p_2^{r_2}$, $\cdots$. (The

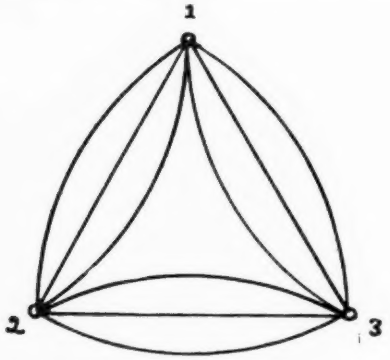
* Cf. Ahrens, Math. Unterhaltungen und Spiele. Leipzig, 1901, Chap. XIII. Encyc. des Sci. Math., Tome I, vol. 3, Fasc. I, p. 72.

proof that type (C) is impossible for degree greater than this value of k is a generalization of the Euler problem of the 36 officers which has not been proved. The simplest case would be to prove that the Euler Square of index 12, 3 is impossible.)

2. **A geometrical interpretation of the Euler Square.** For simplicity we consider first the Euler Square of index 3, 2,

$$\begin{matrix} 1,1 & 2,2 & 3,3, \\ 2,3 & 3,1 & 1,2, \\ 3,2 & 1,3 & 2,1. \end{matrix}$$

The generalization to index n, 2 offers no difficulty. We shall consider the numbers 1, 2, 3 as representing points, and the first column omitting 1, 1 as representing the triangles 1, 2, 3 and 1, 3, 2 where the order of the numbers following 1 is the same as the order of the numbers in the number pairs in the Euler Square. Also in triangle 1, 2, 3 for instance 1 shall be called the first vertex, 2 the second vertex, 3 the third vertex; 1, 2 shall be called the first side, 2, 3 the second side and 3, 1 the third side. To make a diagram in a plane representing the six triangles of this Euler Square, the first sides shall be drawn as straight lines, the second sides as arcs bending outward, the third sides as arcs bending inward; giving the following figure in which segment ij is the same as segment ji only when they are both first sides, second sides, or third sides.



In a more complicated figure the second sides instead of bending outward might be represented by red lines or dotted lines, and the third sides by blue lines or dashed lines.

Evidently then each segment has precisely two regions abutting upon it, for ij is an rth side in but one triangle and ji is an rth side in but one triangle and the two triangles are distinct.

We shall also consider any segment as positively or negatively related to a triangle which it abuts according as the numbers specifying that side

in the notation for the triangle occur in the cyclic order (123) or the cyclic order (132); and a point as positively or negatively related to a segment which it terminates according as it is the first or the second point in the notation for the segment as chosen above.

This Euler Square therefore represents a closed two-sided two-dimensional complex (see "Manifolds of n dimensions," O. Veblen and J. W. Alexander, Annals of Math., vol. 14, p. 164) and the two matrices $A_{0,1}$ and $A_{1,2}$ defining it are as follows, where 1 indicates incidence and positive relation, -1 indicates incidence and negative relation and 0 indicates non-incidence:

		Lines as first sides			Lines as second sides			Lines as third sides		
	Points	1, 2	2, 3	3, 1	1, 2	2, 3	3, 1	1, 2	2, 3	3, 1
	1	1	0	-1	1	0	-1	1	0	-1
$A_{0,1}$:	2	-1	1	0	-1	1	0	-1	1	0
	3	0	-1	1	0	-1	1	0	-1	1

	Triangles	123	132	231	213	312	321
	First Sides						
	12	1	0	0	-1	0	0
	23	0	0	1	0	0	-1
	31	0	-1	0	0	1	0
	Second Sides						
$A_{1,2}$:	12	0	0	0	0	1	-1
	23	1	-1	0	0	0	0
	31	0	0	1	-1	0	0
	Third Sides						
	12	0	-1	1	0	0	0
	23	0	0	0	-1	1	0
	31	1	0	0	0	0	-1

The Euler Square specifies all of the incidence relations of the configuration given by these two matrices in a more compact form.

3. **The Euler Square of index n, 2 for $n \equiv 2 \pmod 4$ is impossible.** From paragraph 2 in the general case the Euler Square of index n, 2 represents a closed two-sided two-dimensional complex with n points, $3n(n-1)/2$ segments and $n(n-1)$ triangular regions.

If the complex is a single two-dimensional circuit (loc. cit., Veblen and Alexander, p. 166), the configuration is a polyhedral region and the $a_0 = n$

points, $a_1 = 3n(n-1)/2$ segments and $a_2 = n(n-1)$ regions satisfy the relation

$$a_0 - a_1 + a_2 = 2 - 2p \tag{1}$$

for some positive integral value of p, in which case p represents the genus of the surface of the polyhedral region. (See Veblen and Young, "Projective Geometry," vol. II, § 188.)

We shall consider the values of p for various Euler squares. If p is not a positive integer no configuration exists of the above type, hence no Euler square exists. If the square of index n, 2 does not exist, then the square of index n, k for $k > 2$ cannot exist; hence we shall first consider squares of index n, 2.

(A) For an Euler Square of index n, 2, if the configuration is a single two-dimensional circuit,

$$a_0 = n, \qquad a_1 = \frac{3}{2}n(n-1), \qquad a_2 = n(n-1).$$

From (1)

$$n - \frac{3}{2}n(n-1) + n(n-1) = 2 - 2p.$$

Then

$$p = 1 + \frac{1}{4}n(n-3).$$

Therefore n must have the form $4k$ or $4k+3$.

(B) If the configuration is separable into m two-dimensional subcircuits, each of the n vertices must occur in the same number m' of circuits. For one of these circuits $a_0 = n_i$, $a_1 = 3k_i/2$, $a_2 = k_i$, therefore

$$n_i - k_i/2 = 2 - 2g_i.$$

Taking the sum of the m equations of this type,

$$m'n - n(n-1)/2 = 2m - 2\textstyle\sum g_i,$$

or

$$n(2m' - n + 1) = 4(m - \textstyle\sum g_i).$$

Therefore n must be a multiple of 4 or $2m' - n + 1$ must be a multiple of 4, in which latter case n must be an odd integer.

In neither case (A) nor (B) can n have the form $4k+2$, therefore the Euler Square is impossible for order $n \equiv 2 \pmod 4$.

If a configuration representing a single circuit determined as above by an Euler Square of index n, 2 be projected on a surface of the same genus so that none of its segments intersect, since at each vertex the same number of segments $3(n-1)$ and the same number of regions $n-1$ meet, there

is determined a regular reticulation of the surface. H. S. White has considered regular reticulations for surfaces of genus $p = 2, 3, \cdots, 9$. Whenever the genus determined by an Euler Square lies in that interval, the corresponding reticulation appears in his list. (H. S. White, "Numerically Regular Reticulations upon Surfaces of Deficiency Higher than One," Bull. Amer. Math. Soc., vol. 3, p. 116, vol. 4, p. 376.)

The following is a table of the genus of the surfaces upon which Euler Squares of order $n = 3, 4, \cdots, 12$ may be developed:

Index	3, 2	4, 2	5, 2	7, 2	7, 2	8, 2	9, 2	11, 2	11, 2	12, 2
Genus	1	2	1 2 circuits	8	1 3 circuits	11	1 4 circuits	23	1 5 circuits	28

4. **Methods of constructing Euler squares.** As the members of the first row are arbitrary subject to the restrictions of the definition, the numbers of the ith k-ad of the first row may all be taken equal to i merely by proper choice of notation. Also since the rows may be permuted the initial members of the first column are taken in the numerical order $1, 2, 3, \cdots$. Furthermore the second k-ad of the first column may be taken in the numerical order $2, 3, 4, \cdots$ since the same permutation may evidently be applied to all the k-ads of an Euler Square.

(*A*) Suppose $n = p$, p a prime > 2. Call G_1 the cyclic group of powers of the substitution $S_1 = (1, 2, 3, \cdots, n)$, and G_2 the cyclic group of the powers of a substitution S_2 of the numbers $2, 3, \cdots, n$, omitting 1, so chosen that it does not send any two numbers to the same two numbers as any substitution of G_1. For $n = 3$ or $n = 5$ there is only one choice for S_2, for $n = 7$ there are 7 choices and the number of choices increases rapidly with n. To construct the Euler Square of index n, $n - 1$ apply the substitutions of G_2 to the $(n - 1)$-ad $2, 3, 4, \cdots, n$ which was chosen as the second member of the first column, to obtain the remaining members of the first column, then apply the substitutions of G_1 to the first column to obtain the other columns.

S_1 and S_2 generate a group G of degree n and order $n(n - 1)$ called the group of the Euler Square. All of the $n(n - 1)$ members of the Euler Square omitting the first row may be obtained by applying the substitutions of G to the $(n - 1)$-ad $2, 3, \cdots, n$. For example for $n = 5$, the Euler Square of index 5, 4 is obtained from $S_1 = (1, 2, 3, 4, 5)$ and $S_2 = (2, 3, 5, 4)$ as follows:

1, 1, 1, 1	2, 2, 2, 2	3, 3, 3, 3	4, 4, 4, 4	5, 5, 5, 5
2, 3, 4, 5	3, 4, 5, 1	4, 5, 1, 2	5, 1, 2, 3	1, 2, 3, 4
3, 5, 2, 4	4, 1, 3, 5	5, 2, 4, 1	1, 3, 5, 2	2, 4, 1, 3
4, 2, 5, 3	5, 3, 1, 4	1, 4, 2, 5	2, 5, 3, 1	3, 1, 4, 2
5, 4, 3, 2	1, 5, 4, 3	2, 1, 5, 4	3, 2, 1, 5	4, 3, 2, 1

In a similar manner an Euler Square can be constructed of index p, $p - 1$ for any prime p.

Remark. A cyclic group of even order has a subgroup of order 2. Therefore any Euler Square of order $2k + 1$ is separable into k Euler Rectangles, because G_2 is a cyclic group of order $2k$ and hence has a sub-group G_3 of order 2. Each Euler Rectangle will give a separate circuit in the configuration, hence an Euler Square of order $2k + 1$ represents k circuits on a surface of genus 1, for

$$a_0 = 2k + 1, \qquad a_1 = 3k(2k + 1), \qquad a_2 = 2k(2k + 1).$$

Therefore from (1) $p = 1$.

For instance, in the square of index 5, 4 above, the first, second and fifth rows form one Euler Rectangle and the first, third and fourth another. In the ith column of an Euler Rectangle the numbers except i occur a number of times equal to the order of the sub-group G_3, hence each number does not appear in every position of the k-ads of a column as is the case in an Euler Square.

(B) Suppose $n = p^r$, p a prime. In this case G_1 cannot be chosen as a cyclic group, but may be chosen as a group of substitutions which are products of p^{r-1} cycles of p numbers each; while G_2 may be chosen as a cyclic group fulfilling the same conditions as in (A), i.e., its substitutions must not transform any two numbers to the same two numbers as any substitution of G_1.

For example, for $n = 2^3$ let G_1 consist of the identity and the following substitutions:

$$\begin{aligned} A &= (12)(34)(56)(78), & B &= (13)(24)(57)(68), \\ C &= (14)(23)(58)(67), & D &= (15)(26)(37)(48), \\ E &= (16)(25)(38)(47), & F &= (17)(28)(35)(46), \\ H &= (18)(27)(36)(45), \end{aligned}$$

and let G_2 be the cyclic group of powers of the substitution $S_2 = (2354786)$; several other choices for S_2 are possible. G_1 and G_2 determine the group of the Euler Square of index 8, 7 by the method given in (A). As a second example, for $n = 3^2$ let G_1 consist of the identity and the following substitutions:

$$\begin{aligned} A &= (123)(468)(597), & B &= (132)(486)(579), \\ C &= (145)(269)(387), & D &= (154)(296)(378), \\ E &= (167)(285)(349), & F &= (176)(258)(394), \\ H &= (189)(247)(365), & J &= (198)(274)(356), \end{aligned}$$

and let G_2 be the cyclic group of powers of the substitution $S_2 = (24693578)$; several other choices for S_2 are possible. G_1 and G_2 generate the Euler Square of index 9, 8.

By the method illustrated in these two examples an Euler Square can be constructed of index p^r, $p^r - 1$ for p any prime.

(*C*) Let $n = 2^r p_1^{r_1} p_2^{r_2} \cdots$; $r, r_1, r_2, \cdots$ positive integers, $r \neq 1$ and $p_1, p_2, \cdots$ distinct odd prime numbers.

Jordan has proved the following theorem (Recherches sur les Substitutions, Liouville Jr. de Math., vol. XVII, 1873, p. 355): "A transitive group of degree n and order $n(n-1)$ whose operations other than the identity displace all or all but one of the symbols can exist only when n is a power of a prime." From this theorem the method used in (*A*) and (*B*) cannot be extended to case (*C*).

For this case we shall use the following method, which is an extension of the method used by G. Tarry (Ahrens, loc. cit.) for degree 2, by combining two Euler Squares of orders a and b to obtain one of order ab; which is similar to the method used for combining two magic squares.

The method may be illustrated as follows, using Euler Squares of indices 5, 3 and 4, 3 to obtain a square of index 20, 3. Given the Euler Square of index 5, 3 as follows:

1, 1, 1	2, 2, 2	3, 3, 3	4, 4, 4	5, 5, 5
2, 3, 4	3, 4, 5	4, 5, 1	5, 1, 2	1, 2, 3
3, 5, 2	4, 1, 3	5, 2, 4	1, 3, 5	2, 4, 1
4, 2, 5	5, 3, 1	1, 4, 2	2, 5, 3	3, 1, 4
5, 4, 3	1, 5, 4	2, 1, 5	3, 2, 1	4, 3, 2

decrease by one all of the numbers of the Euler Square of index 4, 3 given in paragraph 1, giving the following square array:

0, 0, 0	1, 1, 1	2, 2, 2	3, 3, 3
1, 2, 3	0, 3, 2	3, 0, 1	2, 1, 0
2, 3, 1	3, 2, 0	0, 1, 3	1, 0, 2
3, 1, 2	2, 0, 3	1, 3, 0	0, 2, 1,

then replace each triple i, j, k of this array by an entire Euler Square of index 5, 3 obtained from the above Euler Square of index 5, 3 by adding to each of its 25 number triples the numbers $5i$, $5j$, $5k$ respectively. In general by this method we will obtain an Euler Square of index n, k where $k+1$ is the least of the numbers $2^r, p_1^{r_1}, p_2^{r_2}, \cdots$.

The Euler Square of index n, k gives a schedule for a contest between k teams of n members each, where each member is to meet each member of the other teams precisely once, and each member is to participate but once at each field (table, court, etc.) (see E. H. Moore, "Tactical Memoranda, III," Amer. Jr. of Math., vol. XVIII, 1896, p. 264).

GEOMETRIC ASPECTS OF EINSTEIN'S THEORY.

By James Pierpont.

1. **Historical introduction.** Einstein's General Theory of Relativity marks an epoch in physics only comparable with the Principia of Newton. One of its extraordinary features is its intimate interlacement with the foundations of geometry. In the past geometers have imagined different non-euclidean geometries, while the geometry of physicists has remained euclidean. Einstein has broken with this tradition and has shown how the presence of gravitating matter and electricity may determine the character of circumambient space. We wish to show briefly how this has been effected.

To do this we must devote a few words to the origin of his theory in order that the reader may realize how natural, how almost necessary, his generalized theory is. For a long time physicists have tried to develop a satisfactory theory of electro-magnetic phenomena (e.g., light) in moving media. Let us suppose two persons A, A_1 observe a certain phenomenon and that A_1 moves relative to A with a uniform velocity v. A uses a rectangular system S of coördinates x, y, z and a clock to mark the time t. A_1 uses another rectangular system $S_1(x_1, y_1, z_1)$ and a clock time t_1, having the same rate as A's when $v = 0$. Each clock and system of coördinates is at rest relative to its observer. Suppose now that each observer writes down the equations which give an account of the phenomenon. Lorentz showed that a satisfactory theory was obtained if we suppose the equations of A are related to those of A_1 by a certain group of transformations. For simplicity, suppose at a certain instant the axes coincided and that the motion of A_1 is parallel to the x axis. Then these transformations are

$$(1)\qquad x = k(x_1 + vt_1), \qquad y = y_1, \qquad z = z_1, \qquad t = k\left(t_1 + \frac{vx_1}{c^2}\right),$$

where c is the velocity of light in vacuo and $k^2(c^2 - v^2) = c^2$.

A fundamental hypothesis of this theory is that the velocity of light is the same for both observers. Suppose at the time t a light signal has reached the point $P(x, y, z)$, and at the time $t + dt$ its coördinates have changed by dx, dy, dz. Then

$$(2)\qquad c^2 = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2.$$

If P has the coördinates x_1, y_1, z_1 in the system S_1 and dt_1 is the interval of time measured on A_1's clock corresponding to dt, then the velocity being the same,

$$c^2 = \left(\frac{dx_1}{dt_1}\right)^2 + \left(\frac{dy_1}{dt_1}\right)^2 + \left(\frac{dz_1}{dt_1}\right)^2. \tag{3}$$

From (2) and (3) we have

$$c^2dt^2 - dx^2 - dy^2 - dz^2 = 0, \tag{4}$$

$$c^2dt_1^2 - dx_1^2 - dy_1^2 - dz_1^2 = 0. \tag{5}$$

According to the general theory (4) must go over into (5) on applying the transformations (1). This is indeed so.

The next important step we wish to mention in the history of Einstein's theory was taken by Poincaré and Minkowski. They interpreted the quadruple (x, y, z, t) as a point in 4-way space whose metric ds (element of arc) is given by

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2. \tag{6}$$

If we set $ds = 0$, we get (4). It is at this point that quadratic differential forms make their modest entrance on the scene where later they are to play a dominant rôle.

As we have seen, the form (6) remains unaltered for the transformations (1). But this quadratic form remains unaltered by a much wider group. In fact, if we set $c^2t^2 = -w^2$, it goes over, aside from the sign, into

$$dx^2 + dy^2 + dz^2 + dw^2$$

which remains unchanged for all rotations of the (x, y, z, w) axes, i.e., for a group of linear orthogonal transformations. Minkowski, therefore, required that the equations of mathematical physics shall remain unaltered for these transformations, and it became incumbent on the advocates of this theory to find such invariant equations. The execution of this program was practically completed by 1910–11; it finds its best exposition in the book of M. v. Laue, "Das Relativitätsprinzip" (first edition, 1911).

The most salient feature of this theory of relativity is the fact that the equations of transformation involve the time t as well as the space coördinates x, y, z. No one had ever ventured to make so revolutionary a step. That it is possible and often desirable to give the equation of dynamics an invariant form was shown by Lagrange a century and a half ago. We refer to Lagrange's classic equations, e.g.,

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} = 0.$$

and to the invariant equation of Hamilton,

$$\delta \int L dt = 0.$$

We refer also to the researches of Lamé (e.g., Leçons sur les Coordonnées Curvilignes, 1859), to those of Beltrami, and, finally, to Chapters V and VI, "Applications mécanique" and "physiques," in the memorable paper of Ricci and Levi-Civita, "Méthodes de calcul differentiel absolu," in the Mathematische Annalen, vol. 54 (1901).

The foregoing theory depends on the hypothesis that the two observers are moving uniformly relative to each other. Since uniform motion is only an exceptional case, one might urge that a theory which depends on such a limitation must be defective and not worthy of much confidence. Drude voices this opinion in his "Optik" (1912), p. 470, where he says "Allein hieraus ist zu erkennen dass diese 'Theorie' keine physikalische Bedeutung haben kann" and scornfully speaks of it as "dieses Zerrbild."

To turn such objections Einstein sought and found (1913–14) a far broader theory which he and others have developed and which is called the *general* theory of relativity.[1] The older theory outlined above is called the *restricted* theory of relativity.

The new theory may be briefly characterized as follows. When the observer A_1 is moving in a general manner, the relation between the two sets of variables x, y, z, t and x_1, y_1, z_1, t_1 is no longer *linear*. Einstein therefore replaces the quadratic form

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 \tag{7}$$

by the general quadratic form

$$ds^2 = \sum_{i,j} a_{ij}dx_i dx_j, \quad i, j = 1, 2, 3, 4, \; a_{ij} = a_{ji}. \tag{8}$$

To express the equations of physics Einstein has recourse to the *calcul* of Ricci and Levi-Civita mentioned above. The quadratic form (8) is fundamental. From a purely abstract standpoint it furnishes the analytical means of writing down invariant (tensor) equations. On the other hand, by regarding x_1, x_2, x_3, x_4 as coördinates (which in general are not rectangular) of a point in 4-way space, (8) may be regarded as defining the element of arc in this space, i.e., it defines the *metric* in this space since all the metrical properties in the last analysis depend upon (8). The coefficients a_{ij} are only 10 in number since $a_{ij} = a_{ji}$; they are functions of the $x_1, \cdots x_4$. Their determination in any given case depends on the disposition of the gravitating matter and electricity which enter the problem.

For example, in the celebrated problem of the motion of Mercury's perihelion, electrical forces are ignored, the gravitational field is produced by the sun alone, the mass of the planet being neglected in comparison with the sun's. On account of the symmetry of the field it is found that the metric of the surrounding space is given by

$$(9)\qquad ds^2 = -(1 - \mu/r)^{-1}dr^2 - r^2d\varphi^2 - r^2\cos^2\varphi d\theta^2 + (1 - \mu/r)dx_4.$$

Here r, φ, θ are polar coördinates, x_4 is the time coördinate, $\mu = 2km/c = 3\cdot 10^5$ in c.g.s. units, m = mass of sun, $c = 3\cdot 10^{10}$ the velocity of light, and $k = 6, 7\cdot 10^{-8}$ is the constant of gravitation.

For x_4 = constant, $dx_4 = 0$ and (9) reduces to

$$(10)\qquad -ds^2 = (1 - \mu/r)dr^2 + r^2d\varphi^2 + r^2\cos^2\varphi d\theta^2.$$

This defines the metric of the three-dimensional space around the sun. It is not euclidean.

2. **n-way space. Non-euclidean geometry.** These terms are full of mystery to the layman, and it must be confessed that, before the advent of Einstein's theory, few mathematicians and still fewer physicists had more than a bowing acquaintance with these subjects. This is partly due to the unfortunate, one might almost say repulsive, way they have often been presented. To begin with the reader should disabuse himself of the idea that there is an n-way space ($n > 3$) in any such way as we think of our 3-way space. For the purpose of this paper it will be helpful to bear in mind that our geometrical terms are merely geometrical names applied to certain analytical expressions or complexes which have their analogues in our ordinary space. We leave it to the metaphysician to decide whether space is one or many, three or n-dimensional, finite or infinite, etc.

Let $x_1, \cdots x_n$ be n variables, the complex $(x_1, \cdots x_n) = x$ we call a *point* and $x_1, \cdots x_n$ its *coördinates*. The totality of the x's as the coördinates vary form an *n-way space* R_n. Let p be a variable parameter; if the coördinates $x_1, \cdots x_n$ are related by

$$(11)\qquad x_1 = \varphi_1(p), \cdots x_n = \varphi_n(p),$$

the totality or locus of the points x when p ranges over a certain interval is a *curve*. Let p, q be two variable parameters; we say

$$(12)\qquad x_1 = \psi_1(p, q), \cdots x_n = \psi_n(p, q)$$

define a *surface*. A relation $F(x_1, \cdots x_n) = 0$ defines a *hypersurface*. Thus $a_1x_1 + \cdots + a_nx_n = 0$ is a hyperplane.

The metric properties of our space R_n depend on our definition of distance. We say the *distance* between the point x and $x + dx$ is ds where

$$(13)\qquad ds^2 = \sum a_{ij}dx_i dx_j, \quad i, j = 1, \cdots n, \; a_{ij} = a_{ji}.$$

In general the a's are functions of the $x_1, \cdots x_n$.

Example 1. In our ordinary space R_3 (rectangular coördinates)

$$(14)\qquad ds^2 = dx_1^2 + dx_2^2 + dx_3^2.$$

In polar coordinates

$$(15)\qquad ds^2 = dx_1^2 + x_1^2 dx_2^2 + x_1^2 \cos^2 x_2 dx_3^2.$$

Example 2. If R_2 is the surface of a sphere of radius r and x_1, x_2 are the ordinary polar coördinates,

$$(16)\qquad ds^2 = r^2 dx_1^2 + r^2 \cos^2 x_1 dx_2^2.$$

Example 3. In the restricted theory of relativity

$$(17)\qquad ds^2 = c^2 dx_4^2 - dx_1^2 - dx_2^2 - dx_3^2.$$

We call

$$(18)\qquad a = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

the *determinant of the form* (13). For the form (15), for example,

$$a = \begin{vmatrix} 1 & 0 & 0 \\ 0 & x_1^2 & 0 \\ 0 & 0 & x_1^2 \cos^2 x_2 \end{vmatrix} = x_1^4 \cos^2 x_2.$$

Associated with the n^2 coefficients a_{ij} are the quantities

$$(19)\qquad a^{ij} = \frac{A_{ij}}{a},$$

where A_{ij} is the minor of a_{ij} with its proper sign. Since $a_{ij} = a_{ji}$, we have also $a^{ij} = a^{ji}$. A relation of constant use is

$$(20)\qquad \begin{aligned} \sum_\mu a_{\lambda\mu} a^{\nu\mu} &= 1 \quad \text{if} \quad \lambda = \nu. \\ &= 0 \quad \text{if} \quad \lambda \neq \nu, \end{aligned}$$

In fact the well-known relation

$$a_{\lambda 1}A_{\lambda 1} + a_{\lambda 2}A_{\lambda 2} + \cdots + a_{\lambda n}A_{\lambda n} = a,$$

on dividing by a, gives the first, and

$$a_{\lambda 1}A_{\nu 1} + \cdots + a_{\lambda n}A_{\nu n} = 0$$

gives the other.

Example 4. For the form (15) we have $a^{ij} = 0$ if $i \neq j$, and

$$a^{11} = \frac{x_1^2 \cdot x_1^2 \cos^2 x_2}{a} = 1, \qquad a^{22} = \frac{x_1^2 \cos^2 x_2}{a} = \frac{1}{x_1^2}, \qquad a^{33} = \frac{1}{x_1^2 \cos^2 x_2}.$$

Another metric notion of great importance is the *angle* θ between two lines, or in general between two curves meeting at a point x. We define this by

$$\cos\theta = a_{ij}\frac{dx_i}{ds}\frac{\delta x_j}{\delta s}, \tag{21}$$

where ds, δs are the elements of arc along the two curves and dx_i, δx_j are the coördinate differences of the extremities of these arcs.

Example 5. Using the form (14),

$$\begin{aligned}\cos\theta &= \frac{dx_1}{ds}\frac{\delta x_1}{\delta s} + \frac{dx_2}{ds}\frac{\delta x_2}{\delta s} + \frac{dx_3}{ds}\frac{\delta x_3}{\delta s}\\ &= l\lambda + m\mu + n\nu,\end{aligned} \tag{21a}$$

where l, m, n and λ, μ, ν are the direction cosines of the two curves at the point x.

When $\cos\theta = 0$, we say the curves *meet at right angles* or *orthogonally*.

When p varies from $p = \alpha$ to $p = \beta$, $\alpha < \beta$, the *length of the arc* on the curve (11) is defined to be

$$s = \int_\alpha^\beta \sqrt{\sum a_{ij}\frac{dx_i}{dp}\frac{dx_j}{dp}}\,dp. \tag{22}$$

We need one other metric notion, that of area for an R_2 and of volume for an R_n, $n > 2$. Calling this V whether $n = 2$ or $n > 2$, we define

$$V = \int \sqrt{|a|}\, dx_1 \cdots dx_n. \tag{23}$$

Example 6. Using the metric of example 2 we have $\sqrt{a} = r^2 \cos x_1$, hence for the whole sphere

$$V = \int_0^{2\pi} dx_2 \int_{-(\pi/2)}^{\pi/2} r^2 \cos x_1 dx_1 = 4\pi r^2.$$

It is important to note that the expressions (21), (22) defining angle and volume are invariant under any transformation. To illustrate what this means, suppose we transform the variables $x_1, \cdots x_n$ to $u_1, \cdots u_p$ whereby ds^2 as given by (13) goes over into

$$d\sigma^2 = \sum_{\alpha,\beta} b_{\alpha\beta}du_\alpha du_\beta, \qquad \alpha, \beta = 1, 2, \cdots p.$$

If we make this transformation in (21), we find it goes over into

$$\cos\theta = \sum b_{\alpha\beta}\frac{du_\alpha}{d\sigma}\frac{\delta u_\beta}{\delta\sigma}, \tag{24}$$

i.e., (24) is the same function of the new letters as (21) is of the old. If, in particular, $d\sigma^2 = du_1^2 + du_2^2 + \cdots + du_p^2$,

$$\cos\theta = \frac{du_1}{d\sigma}\frac{\delta u_1}{\delta\sigma} + \frac{du_2}{d\sigma}\frac{\delta u_2}{\delta\sigma} + \cdots + \frac{du_p}{d\sigma}\frac{\delta u_p}{\delta\sigma}.$$

If $n = 3$, this reduces to (21a), i.e., the angle θ is the same as in the corresponding three-dimensional ordinary space.

3. **Geodesics.** These curves take the place of right lines, whence their importance in non-euclidean geometry. To better understand their definition, which will be given presently, let us consider the integral

$$(25) \qquad A = \int_\alpha^\beta \varphi(x, y, z, u, v)dp$$

taken over the curve C whose equations are $x = x(p)$, $y = y(p)$, $z = z(p)$. Here u, v are functions of p, x, y, z and their derivatives. Let us in this integral replace x, y, z by $\bar{x} = x + \delta x$, $\bar{y} = y + \delta y$, $\bar{z} = z + \delta z$. Geometrically speaking we replace C by an adjacent curve having however the same endpoints. At the same time u becomes $u + \delta u$, v becomes $v + \delta v$, while φ becomes $\bar{\varphi} = \varphi(x + \delta x, \cdots u + \delta u, v + \delta v)$, which, developed by Taylor's theorem, gives

$$\delta\varphi = \bar{\varphi} - \varphi = \frac{\partial\varphi}{\partial x}\delta x + \cdots + \frac{\partial\varphi}{\partial v}\delta v,$$

neglecting small quantities beyond the first order. Then

$$\delta A = \bar{A} - A = \int_\alpha^\beta \bar{\varphi}dp - \int_\alpha^\beta \varphi dp = \int_\alpha^\beta \delta\varphi dp.$$

When the original curve C is such that $\delta A = 0$, we say the curve renders the integral (25) *stationary*. Ordinarily it corresponds to a maximum or minimum value of A.

Let us apply these considerations to the integral

$$s = \int_\alpha^\beta \sqrt{\Sigma a_{ij}\frac{dx_i}{dp}\frac{dx_j}{dp}}\,dp = \int ds,$$

which gives the length of an arc of the curve (11). If this curve is such that

$$(26) \qquad \delta\int ds = 0,$$

we say that it is a *geodesic*, ordinarily it is the shortest curve between the two fixed points $p = \alpha$, $p = \beta$. The variational equation (26) leads easily to the n equations

$$(27) \qquad \sum_{i,j}\frac{dx_i}{ds}\frac{dx_j}{ds}\frac{\partial a_{ij}}{dx_k} - 2\sum_1\frac{d}{ds}\left(a_{ik}\frac{dx_i}{ds}\right) = 0, \qquad k = 1, 2, \cdots n.$$

Example 7. In case $n = 2$, these equations become, on setting $x_1 = u$, $x_2 = v$,

$$(28)\quad \begin{aligned} 2\frac{d}{ds}\left[a_{11}\frac{du}{ds}+a_{12}\frac{dv}{ds}\right] &= \frac{\partial a_{11}}{\partial u}\left(\frac{du}{ds}\right)^2 + 2\frac{\partial a_{12}}{\partial u}\frac{du}{ds}\frac{dv}{ds} + \frac{\partial a_{22}}{\partial u}\left(\frac{dv}{ds}\right)^2, \\ 2\frac{d}{ds}\left[a_{21}\frac{du}{ds}+a_{22}\frac{dv}{ds}\right] &= \frac{\partial a_{11}}{\partial v}\left(\frac{du}{ds}\right)^2 + 2\frac{\partial a_{12}}{\partial v}\frac{du}{ds}\frac{dv}{ds} + \frac{\partial a_{22}}{\partial v}\left(\frac{dv}{ds}\right)^2. \end{aligned}$$

It should be noticed that, if one of these equations is satisfied, the other is also.

Example 8. Let us consider as a special case a surface of revolution,

$$x = v\cos u, \qquad y = v\sin v, \qquad z = \varphi(v),$$

using the ordinary definition of $ds^2 = dx^2 + dy^2 + dz^2$. Then for an arc on this surface $ds^2 = v^2du^2 + (1+\psi)dv^2$, where $\psi = (d\varphi/dv)^2$.

From (28) we can show at once that the meridians u = constant are geodesics on this surface. For along a meridian $du/ds = 0$, also $a_{12} = 0$ and $\partial a_{22}/\partial u = 0$. Hence both sides of the first equation of (28) are identically zero. Thus u = constant is a solution of our differential equations. The parallels, z = constant, are the orthogonal trajectories to these geodesics.

Reverting to the general case we notice that the equations (27) involve the second derivatives of the coördinates x_i. In order to solve the equations with respect to these quantities we introduce the *symbols of Christoffel* which pervade Einstein's theory. They are

$$(29)\quad \begin{bmatrix}\alpha\ \beta\\ k\end{bmatrix} = \frac{1}{2}\left(\frac{\partial a_{\alpha k}}{\partial x_\beta} + \frac{\partial a_{\beta k}}{\partial x_\alpha} - \frac{\partial a_{\alpha\beta}}{\partial x_k}\right) \qquad \alpha,\beta,k = 1,2,\cdots n,$$

$$(30)\quad \begin{Bmatrix}\alpha\ \beta\\ \lambda\end{Bmatrix} = \sum_k a^{\lambda k}\begin{bmatrix}\alpha\ \beta\\ k\end{bmatrix} \qquad k = 1,2,\cdots n.$$

It is important to notice that they are symmetric in α, β.

Example 9. $ds^2 = x_2^2dx_1^2 + dx_2^2$ (element of arc in polar coördinates). Here $a_{11} = x_2^2$, $a_{12} = a_{21} = 0$, $a_{22} = 1$, $a = x_2^2$, $a^{11} = 1/x_2^2$, $a^{12} = a^{21} = 0$, $a^{22} = 1$.

$$\begin{bmatrix}1\ 1\\ 1\end{bmatrix} = \frac{1}{2}\frac{\partial a_{11}}{\partial x_1} = 0, \qquad \begin{bmatrix}1\ 2\\ 1\end{bmatrix} = \frac{1}{2}\frac{\partial a_{11}}{\partial x_2} = x_2,$$

$$\begin{bmatrix}2\ 2\\ 1\end{bmatrix} = \frac{\partial a_{12}}{\partial x_2} - \frac{1}{2}\frac{\partial a_{22}}{\partial x_1} = 0,$$

$$\begin{bmatrix}1\ 1\\ 2\end{bmatrix} = -x_2, \qquad \begin{bmatrix}1\ 2\\ 2\end{bmatrix} = 0, \qquad \begin{bmatrix}2\ 2\\ 2\end{bmatrix} = 0,$$

$$\begin{Bmatrix}1\ 1\\ 1\end{Bmatrix} = a^{11}\begin{bmatrix}1\ 1\\ 1\end{bmatrix} + a^{21}\begin{bmatrix}1\ 1\\ 2\end{bmatrix} = 0,$$

$$\begin{Bmatrix}1\ 2\\ 1\end{Bmatrix} = a^{11}\begin{bmatrix}1\ 2\\ 1\end{bmatrix} + a^{21}\begin{bmatrix}1\ 2\\ 2\end{bmatrix} = \frac{1}{x_2},$$

$$\begin{Bmatrix} 2\ 2 \\ 1 \end{Bmatrix} = 0, \qquad \begin{Bmatrix} 1\ 1 \\ 2 \end{Bmatrix} = -x_2, \begin{Bmatrix} 1\ 2 \\ 2 \end{Bmatrix} = 0, \qquad \begin{Bmatrix} 2\ 2 \\ 2 \end{Bmatrix} = 0.$$

In terms of the symbols $\{{}^{i}{}_{\lambda}{}^{j}\}$ we may write the n equations (27)

$$\frac{d^2x_\lambda}{ds^2} + \sum_{ij} \begin{Bmatrix} i\ j \\ \lambda \end{Bmatrix} \frac{dx_i}{ds}\frac{dx_j}{ds} = 0, \qquad \lambda = 1, 2, \cdots n. \tag{31}$$

These are the equations of a geodesic employed by Einstein. As he supposes that a body moving freely in a gravitational field describes a geodesic, these are the equations of motion of this body as, for example, Mercury about the sun (here $n = 4$). They depend entirely upon ds, that is, the metric of the surrounding space.

Example 10. Let $ds^2 = \sum a_{ij}dx_idx_j$, the coefficients a_{ij} being constant. From (29) we see all the symbols $[{}^{\alpha}{}_{k}{}^{\beta}] = 0$ as the a's are constant. Thus by (30) all the $\{{}^{\alpha}{}_{\lambda}{}^{\beta}\} = 0$. Hence (31) reduces to the n equations $d^2x_\lambda/ds^2 = 0$, $\lambda = 1, 2, \cdots n$. Integrating we get $x_\lambda = A_\lambda s + B_\lambda$, A_λ and B_λ being constants. These are the equations of a right line, the parameter being s. Thus when the coefficients a_{ij} which define ds^2 are constants, the geodesics in this space are right lines. This is the case in the restricted theory of relativity (6), since there the velocity c of light in vacuo is constant.

4. **Elliptic space.** As we shall see, Einstein assumes that our space is not infinite in extent. It has a definite volume like a sphere, viz., $V = \pi^2R^3$, where R has the approximate value $R = 9\cdot 10^{11}$ *orbrads*, this unit being the mean distance of the earth from the sun, i.e., 1 orbrad = 150 million kilometers. All geodesics (pseudo right lines) are closed curves and have the length πR. Thus, were it not for the absorption of light in traversing such enormous distances, to the sun should correspond another sun, a sort of anti-sun, in the opposite direction. Such a space may seem preposterous to the naïve mind, but so did the existence of people living at the antipodes a few hundred years ago. The first to study an elliptic space R_n, $n > 2$, was Riemann; a 2-way space of this type has been known since the days of the Greeks, it is the surface of a sphere.

Without going into details let us show how the properties of this space may be easily deduced. To this end we take a set of rectangular axes in the euclidean plane and define the position of a point by the coördinates x, y measured in the ordinary way. The distance ds between the point x, y and the point $x + dx$, $y + dy$ we define by

$$ds^2 = \frac{dx^2 + dy^2}{\left[1 + \dfrac{x^2 + y^2}{4R^2}\right]^2} = \frac{16R^4}{\lambda^2}(dx^2 + dy^2), \qquad \lambda = x^2 + y^2 + 4R^2. \tag{32}$$

The metric of this R_2 is not euclidean; but we may refer it to a euclidean R_3 as follows. Set

$$(33) \qquad u = \frac{4R^2x}{\lambda}, \qquad v = \frac{4R^2y}{\lambda}, \qquad w = \frac{8R^3}{\lambda} - R.$$

Then

$$du = \frac{4R^2}{\lambda^2}(\lambda dx - xd\lambda), \qquad dv = \frac{4R^2}{\lambda^2}(\lambda dy - yd\lambda), \qquad dw = -\frac{8R^3d\lambda}{\lambda^2},$$

from which follows

$$(34) \qquad ds^2 = du^2 + dv^2 + dw^2.$$

Thus to each point x, y in the elliptic plane corresponds a point u, v, w in our ordinary three-dimensional space. This illustrates the important theorem: *If the metric of an R_n is defined by*

$$ds^2 = \sum a_{ij}dx_idx_j, \qquad i, j = 1, 2, \cdots n,$$

we may choose $m + n$ new variables $u_1, \cdots u_{m+n}$ such that

$$ds^2 = du_1^2 + du_2^2 + \cdots + du_{m+n}^2;$$

moreover $m \leqq n(n-1)/2$.

Thus we may regard the n-way space R_n as embedded in an $(m+n)$-way euclidean space. From (33) we find that

$$(35) \qquad u^2 + v^2 + w^2 = R^2.$$

Thus when x, y ranges over the elliptic plane, the point u, v, w ranges over a sphere.

Let us now see what conclusions we can draw relative to the geometry of this plane R_2. In the first place we find

$$x = \frac{2Ru}{R+w}, \qquad y = \frac{2Rv}{R+w}.$$

To each point u, v, w corresponds a single point x, y with one exception, viz., when $R + w = 0$. But then $u = v = 0$, as (35) shows. The correspondence between R_2 and R_3 is thus 1 to 1 with this one exception.

The geodesics or, as we shall call them, the *pseudo right lines*, are determined by

$$\delta \int ds = 0,$$

where ds is defined by (32). If we change to the u, v, w variables, ds is defined by (34) subject, however, to the relation (35). Thus, to pseudo right lines correspond geodesics on the sphere (35), i.e., to great circles on this sphere. From this we have:

(i) All pseudo right lines in this R_2 are closed curves.

(ii) Their length is $2\pi R$.

(iii) Two pseudo right lines meet in two points. Hence

(iv) There are no parallels.

Let C_1, C_2 be two curves making an angle θ with each other as defined by (21). On changing to the u, v, w variables these curves go over into two curves Γ_1, Γ_2 on the sphere, and $ds^2 = du^2 + dv^2 + dw^2$. But in this case we saw that θ is the angle between Γ_1 and Γ_2 in the ordinary way. Hence we have

(v) The trigonometry of our R_2 is the trigonometry on a sphere of radius R. The sum of the angles of a triangle formed by three pseudo right lines is always greater than 180°.

Since all great circles on a sphere perpendicular to a given great circle meet at a point, viz., the pole of this circle, and hence have the length $\pi R/2$, we have

(vi) All pseudo right lines in the elliptic plane perpendicular to a given pseudo right line meet at a point and have a common length $\pi R/2$.

We have so far made no attempt to visualize the pseudo right lines in the elliptic plane. It is easy to do this; for on the sphere they correspond to great circles. Let one of these great circles lie in the plane $Au + Bv + Cw = 0$. Replacing u, v, w by their values in (33) we get

$$ARx + BRy - (x^2 + y^2 + 4R^2) + 8R^2C = 0,$$

the equation of a circle in the (euclidean) x, y plane. In particular, the pseudo right line corresponding to the equation $w = 0$ is the circle

$$x^2 + y^2 = 4R^2, \tag{36}$$

which we call the *fundamental circle*. Since all great circles cut the equator in diametrically opposite points, we see that all pseudo right lines cut the fundamental circle in such points. Conversely, such circles are pseudo right lines in our elliptic geometry.

The geometry so far developed differs from plane euclidean geometry therein that its pseudo right lines cut a given pseudo right line twice. We may, if we like, agree to regard all points of the x, y plane outside the fundamental circle as non-existent as far as our elliptic geometry is concerned. Also we shall assume that diametrically opposite points of this circle are identical. In this case two pseudo right lines cut once only and they all have the common length πR.

Let us turn now to elliptic space; as the work is entirely analogous, we may be more brief. We start with a set of rectangular coördinates and define a point by the coördinates x, y, z measured in the ordinary way. We define the metric by

$$(37)\quad ds^2 = \frac{dx^2 + dy^2 + dz^2}{\left[1 + \frac{1}{4R^2}(x^2 + y^2 + z^2)\right]^2} = \frac{16R^2(dx^2 + dy^2 + dz^2)}{[x^2 + y^2 + z^2 + 4R^2]^2}.$$

As before we set $\lambda = x^2 + y^2 + z^2 + 4R^2$ and

$$(38)\quad u_1 = \frac{4R^2x}{\lambda}, \qquad u_2 = \frac{4R^2y}{\lambda}, \qquad u_3 = \frac{4R^2z}{\lambda}, \qquad u_4 = \frac{8R^3}{\lambda} - R,$$

and find again that

$$ds^2 = du_1^2 + du_2^2 + du_3^2 + du_4^2.$$

In an entirely analogous manner we find that geodesics in this space or, as we prefer to call them, pseudo right lines cut the fundamental sphere

$$(39)\quad x^2 + y^2 + z^2 = 4R^2$$

in diametrically opposite points.

The analogue of the euclidean plane is a sphere cutting the fundamental sphere along a great circle. We call it a pseudo plane. As before we have two geometries according as we regard opposite ends of a diameter of the fundamental sphere (39) as identical or not. In the former case points outside of (39) are non-existent. Einstein in his cosmological considerations prefers this type of geometry. In this sphere we have:

(i) All pseudo right lines have the length πR and are closed curves.
(ii) These lines cut once only.
(iii) There are no parallels.
(iv) Two points determine a pseudo right line.
(v) Three points determine a pseudo plane.

According to this geometry the whole physical universe lies within the fundamental sphere. Let us find the volume. By (23)

$$(40)\quad V = \int \sqrt{a}\, dxdydz.$$

From (37) we have, setting $\alpha^{-1} = 1 + \rho^2(x^2 + y^2 + z^2)$, $\rho = 1/2R$,

$$a = \begin{vmatrix} \alpha^2 & 0 & 0 \\ 0 & \alpha^2 & 0 \\ 0 & 0 & \alpha^2 \end{vmatrix} = \alpha^6.$$

Let us change the variables in the integral (40), setting

$$x = r\cos\theta\cos\varphi, \qquad y = r\cos\theta\sin\varphi, \qquad z = r\sin\theta.$$

Then

$$V = \int J\, \frac{drd\theta d\varphi}{R},$$

where

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} & \frac{\partial z}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} & \frac{\partial z}{\partial \theta} \\ \frac{\partial x}{\partial \varphi} & \frac{\partial y}{\partial \varphi} & \frac{\partial z}{\partial \varphi} \end{vmatrix}.$$

Thus

$$(41)\quad V = \int_0^{2R} \frac{r^2dr}{(1+\rho^2r^2)^3} \int_{-\pi/2}^{\pi/2} \cos\theta d\theta \int_0^{2\pi} d\varphi = 4\pi \int_0^{2R} \frac{r^2dr}{(1+\rho^2r^2)^3} = \pi^2R^3.$$

As we have remarked and as we shall see later, an approximate value of R is $9\cdot 10^{11}$ times the mean distance of the earth from the sun.

5. **Curvature.** The metric properties of a given R_n depend, as we have seen, on the definition of distance between two nearby points, i.e., on the quadratic form

$$(42)\qquad ds^2 = \sum a_{ij}dx_idx_j.$$

Another space R_n', whose metric is defined by quite a different expression

$$(43)\qquad d\sigma^2 = \sum b_{ij}du_idu_j,$$

may have essentially the same geometry. For example, in R_2 let $ds^2 = dx_1^2 + dx_2^2$, and in R_2' let $d\sigma^2 = u_2^2du_1^2 + du_2^2$. If we set

$$(44)\qquad x_1 = u_2 \cos u_1, \qquad x_2 = u_2 \sin u_1,$$

we find $ds^2 = d\sigma^2$. The relations (44) enable us to establish a 1 to 1 correspondence between the points of R_2 and R_2'. Since $ds = d\sigma$, corresponding arcs have the same length and corresponding angles are equal. Hence their metrical properties are the same.

We are therefore led to ask when is it possible, by a suitable change of variables, to transform (42) into (43), and conversely. Without answering this with entire generality we may give a partial answer sufficient for our purpose. To this end we introduce the *symbols of Riemann*.

$$(45)\quad \{\alpha\beta, \lambda\mu\} = \frac{\partial \begin{Bmatrix} \alpha & \lambda \\ & \beta \end{Bmatrix}}{\partial x_\mu} - \frac{\partial \begin{Bmatrix} \alpha & \mu \\ & \beta \end{Bmatrix}}{\partial x_\lambda} + \sum \left[\begin{Bmatrix} \alpha & \lambda \\ & k \end{Bmatrix}\begin{Bmatrix} k & \mu \\ & \beta \end{Bmatrix} - \begin{Bmatrix} \alpha & \mu \\ & k \end{Bmatrix}\begin{Bmatrix} k & \lambda \\ & \beta \end{Bmatrix} \right], \quad k = 1, 2, \cdots n,$$

and

$$(46)\qquad (\alpha\,\gamma, \lambda\,\mu) = \sum_\beta a_{\beta\gamma}\{\alpha\,\beta, \lambda\,\mu\}.$$

As in the case of the Christoffel symbols we have

$$(47)\qquad \{\alpha, \beta, \lambda, \mu\} = \sum_\gamma a^{\beta\gamma}(\alpha\,\gamma, \lambda\,\mu).$$

By means of (46) we may separate out an important class of n-way spaces called *spaces of constant curvature.** We say R_n has constant curvature k when for all $\alpha, \beta, \lambda, \mu = 1, 2, \cdots n$

$$(\alpha\,\beta, \lambda\,\mu) = k \begin{vmatrix} a_{\alpha\lambda} & a_{\alpha\mu} \\ a_{\beta\lambda} & a_{\beta\mu} \end{vmatrix}. \tag{48}$$

We may now state the important theorem: *If R_n, R_n' are two spaces of the same constant curvature k, we may transform* (42) *into* (43) *by a suitable change of variable, and conversely; that is, $ds = d\sigma$.* The metric properties of the two spaces are the same, at least for *sufficiently restricted regions.*

Riemann showed that for spaces of constant curvature k the element of arc may be defined by

* As the term *curvature* figures so largely in Einstein's theory and quite wrong ideas are current in some quarters, a few additional words of explanation may be acceptable. In ordinary space the curvature of a surface S at a point x is defined by

$$k = \frac{1}{R_1R_2}, \tag{a}$$

where R_1, R_2 are the greatest and least radii of curvature of the normal sections of S at x. Gauss made the extraordinary discovery that k remains invariant under all transformations of the variables. We find in fact that, if the metric of S is given by

$$ds^2 = a_{11}du_1^2 + 2a_{12}du_1du_2 + a_{22}du_2^2, \tag{b}$$

then

$$k = \frac{(12, 12)}{a}. \tag{c}$$

Suppose now the surface S, lying in an n-way space R_n, is defined by $x_1 = x_1(u_1, u_2), \cdots x_n = x_n(u_1, u_2)$. If the metric of R_n is defined by (13), the element of arc $d\sigma$ on S is given by

$$d\sigma^2 = \sum_{i,j} a_{ij} \sum_k \frac{\partial x_i}{\partial u_k} du_k \sum_l \frac{\partial x_j}{\partial u_l} du_l, \qquad i, j = 1, 2, \cdots n; \quad k, l = 1, 2,$$

or

$$d\sigma^2 = g_{11}du_1^2 + 2g_{12}du_1du_2 + g_{22}du_2^2. \tag{d}$$

The curvature of S at a point x is now *defined* by

$$k = \frac{(12, 12)_a}{g}, \tag{e}$$

where g is the determinant of the quadratic form (d) and $(12, 12)_a$ is the Riemannian symbol (46) relative to this form. We see this definition is merely an extension of (c) from 3 to n-way space. But, whereas (c) has the geometric interpretation (a), the definition (e) has not, it is merely an analytic generalization. The reader should not undervalue it on that score; its importance is fundamental.

Let us now consider a curve C. The n quantities $\eta_1 = dx_1/ds, \cdots \eta_n = dx_n/ds$ are called the *directional parameters* of C at a given point x. Through any point x of our R_n there passes a geodesic having a given $\eta = (\eta_1, \eta_2, \cdots \eta_n)$. If $\eta' = (\eta_1', \cdots \eta_n')$ is another set of parameters at x, $g\eta + g'\eta$ will denote a pencil of these parameters, g, g' being variables. To this pencil corresponds a pencil of geodesics through x having $g\eta + g'\eta'$ as directional parameters. These geodesics constitute a surface G in R_n on which an element of arc $d\sigma^2$ has the form (d). The curvature k of G at the point x is given by (e). Suppose now that k has the same constant value however the pencil (η, η') is oriented about x, we say R_n is a *space of constant curvature* k.

$$ds^2 = \frac{dx_1^2 + dx_2^2 + \cdots + dx_n^2}{[1 + \frac{1}{4}k^2(x_1^2 + x_2^2 + \cdots + x_n^2)]^2}. \tag{49}$$

When $k = 0$, this reduces to $ds^2 = dx_1^2 + dx_2^2 + \cdots dx_n^2$. In this case we have seen that the geodesics are right lines, the $x_1, \cdots x_n$ being referred to a rectangular coördinate system (for clearness the reader may suppose $n = 3$). We therefore regard k as a measure of the departure of the space R_n defined by (42) from euclidean space. Thus we saw in the elliptic space R_3 that the geodesics, instead of being straight lines, are arcs of circles. Here $k = 1/R^2$, as is seen by comparing (37) and (49). The smaller k is, the more nearly these geodesics or pseudo right lines approach straight lines in euclidean space.

Example 11. Let us see if the R_2 whose metric is defined by

$$ds^2 = c^2 \cos^2 x_2 dx_1^2 + c^2 dx_2^2 \tag{50}$$

has constant curvature. Here

$$a_{11} = c^2 \cos^2 x_2, \qquad a_{12} = a_{21} = 0, \qquad a_{22} = c^2, \qquad a = c^4 \cos^2 x_2,$$

$$a^{11} = \frac{1}{a_{11}}, \qquad a^{12} = a^{21} = 0, \qquad a^{22} = \frac{1}{a_{22}},$$

$$\begin{bmatrix} 1\ 1 \\ 1 \end{bmatrix} = 0, \qquad \begin{bmatrix} 1\ 2 \\ 1 \end{bmatrix} = -c^2 \sin x_2 \cos x_2, \qquad \begin{bmatrix} 2\ 2 \\ 1 \end{bmatrix} = 0,$$

$$\begin{bmatrix} 1\ 1 \\ 2 \end{bmatrix} = c^2 \sin x_2 \cos x_2, \qquad \begin{bmatrix} 1\ 2 \\ 2 \end{bmatrix} = 0, \qquad \begin{bmatrix} 2\ 2 \\ 2 \end{bmatrix} = 0,$$

$$\begin{Bmatrix} 1\ 1 \\ 1 \end{Bmatrix} = 0, \qquad \begin{Bmatrix} 1\ 2 \\ 1 \end{Bmatrix} = -\tan x_2, \qquad \begin{Bmatrix} 2\ 2 \\ 1 \end{Bmatrix} = 0,$$

$$\begin{Bmatrix} 1\ 1 \\ 2 \end{Bmatrix} = \sin x_2 \cos x_2, \qquad \begin{Bmatrix} 1\ 2 \\ 2 \end{Bmatrix} = 0, \qquad \begin{Bmatrix} 2\ 2 \\ 2 \end{Bmatrix} = 0,$$

$$(1\,2,\ 1\,2) = a_{12}\{1\,1,\ 1\,2\} + a_{22}\{1\,2,\ 1\,2\} = c^2\{1\,2,\ 1\,2\},$$

$$\{1\,2,\ 1\,2\} = \frac{\partial}{\partial x_2}\begin{Bmatrix} 1\ 1 \\ 2 \end{Bmatrix} - \begin{Bmatrix} 1\ 2 \\ 1 \end{Bmatrix}\begin{Bmatrix} 1\ 1 \\ 2 \end{Bmatrix}$$

$$= -\sin^2 x_2 + \cos^2 x_2 + \tan x_2 \sin x_2 \cos x_2 = \cos^2 x_2,$$

$$\therefore \quad (1\,2,\ 1\,2) = c^2 \cos^2 x_2.$$

From (48)

$$(1\,2,\ 1\,2) = k \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = ak, \qquad \therefore\ k = \frac{1}{c^2}.$$

By using the fact that $(\alpha\,\beta,\ \lambda\,\mu)$ changes its sign when we interchange α, β or λ, μ, and hence is zero when $\alpha = \beta$ or $\lambda = \mu$, we find that all the $2^4 = 16$ symbols $(\alpha\,\beta,\ \lambda\,\mu)$ placed in (48) either give $k = 1/c^2$ or $0 = k \cdot 0$. Thus the 16 relations (48) are satisfied by this value of k. Hence R_2 as defined by (50) is a 2-way space of constant curvature $k = 1/c^2$. In

fact if we regard x_1, x_2 as longitude and latitude, (50) defines ds on a sphere of radius c.

Example 12.

(51) $$ds^2 = c_1dx_1^2 + c_2dx_2^2 + \cdots + c_ndx_n^2,$$

where the coefficients are constants. Here all the $[{}^{\alpha}{}_{k}{}^{\beta}] = 0$ since the a_{ij} are constants, a_{ij} being zero if $i \neq j$. Hence all the $\{{}^{\alpha}{}_{k}{}^{\beta}\} = 0$. Thus all the $\{\alpha\,\beta, \lambda\,\mu\} = 0$, and hence finally all the $(\alpha\,\beta, \lambda\,\mu) = 0$. Thus the n^4 equations (48) are satisfied by $k = 0$. The curvature of the space defined by (50) is therefore 0. A special case of (51) is

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2,$$

which defines the metric of the 4-way space of the restricted theory of relativity. Although we call a space for which $k = 0$ euclidean, the reader should note that such a space may possess pseudo lines of null length, i.e., lines for which $ds = 0$. If we set $ds = 0$ in the last equation, we get

$$c^2dt^2 - dx^2 - dy^2 - dz^2 = 0,$$

which is (4). Thus the path of a ray of light is a null line in the restricted theory of relativity.

6. **Tensors.** To form invariant differential equations expressing the laws of physics, Einstein found ready at hand a calculus which seems almost created for his needs. This is the *calcul différentiel absolu* of Ricci and Levi-Civita already referred to. We think a better name is *tensor analysis*. To give the reader a concrete example of a tensor, in fact one of the most important tensors, let us see how the n^2 coefficients a_{ij} in

(52) $$ds^2 = \sum_{i,j} a_{ij}dx_idx_j, \qquad i, j = 1, 2, \cdots n,$$

behave when we replace the variables $x_1, \cdots x_n$ by n new variables $u_1, \cdots u_n$. Since

$$dx_i = \sum_{\lambda} \frac{\partial x_i}{\partial u_\lambda} du_\lambda, \qquad \lambda = 1, 2, \cdots n,$$

(52) becomes

$$ds^2 = \sum_{i,j} a_{ij} \sum_{\lambda} \frac{\partial x_i}{\partial u_\lambda} du_\lambda \sum_{\mu} \frac{\partial x_j}{\partial u_\mu} dx_\mu = \sum_{\lambda,\mu} du_\lambda du_\mu \sum_{i,j} a_{ij} \frac{\partial x_i}{\partial u_\lambda}\frac{\partial x_j}{\partial u_\mu}.$$

Hence

$$ds^2 = \sum \bar{a}_{\lambda\mu} du_\lambda du_\mu, \qquad \lambda, \mu = 1, 2, \cdots n,$$

where

(53) $$\bar{a}_{\lambda\mu} = \sum_{i,j} \frac{\partial x_i}{\partial u_\lambda}\frac{\partial x_j}{\partial u_\mu} a_{ij}.$$

Let us generalize and say that the n^2 functions A_{ij} of $x_1, \cdots x_n$ form

a covariant tensor of order 2 if, on changing the variables to $u_1, \cdots u_n$, the transformed $\bar{A}_{\lambda\mu}$ are related to the old A_{ij} by

$$\bar{A}_{\lambda\mu} = \sum_{i,j} \frac{\partial x_i}{\partial u_\lambda} \frac{\partial x_j}{\partial u_\mu} A_{ij}. \tag{54}$$

The individual A_{ij} are called the components of this tensor. From this we see the n^2 coefficients a_{ij} in (52) form a covariant tensor of order 2.

We may generalize (54) as follows. Suppose we have n^k functions $A_{\alpha\beta\cdots\kappa}$ of $x_1, \cdots x_n$ which are transformed according to

$$\bar{A}_{\lambda\mu\cdots\omega} = \sum \frac{\partial x_\alpha}{\partial u_\lambda} \frac{\partial x_\beta}{\partial u_\mu} \cdots \frac{\partial x_\kappa}{\partial u_\omega} A_{\alpha\beta\cdots\kappa}, \tag{55}$$

the summation extending over the k indices $\alpha, \beta, \cdots \kappa$ from 1 to n. We say these n^k functions form a *covariant tensor of order* k.

Example 13. The n^4 symbols of Riemann $(\alpha\, \beta, \gamma\, \delta)$ are transformed in this way. In fact, if we set

$$G_{\alpha\beta\gamma\delta} = (\alpha\, \beta, \gamma\, \delta), \tag{56}$$

we find by a reasoning too long to give here that, on changing variables,

$$\bar{G}_{\lambda\mu\nu\omega} = \sum_{\alpha,\beta,\gamma,\delta} \frac{\partial x_\alpha}{\partial u_\lambda} \frac{\partial x_\beta}{\partial u_\mu} \frac{\partial x_\gamma}{\partial u_\nu} \frac{\partial x_\delta}{\partial u_\omega} G_{\alpha\beta\gamma\delta}, \quad \alpha, \beta, \gamma, \delta = 1, 2, \cdots n. \tag{57}$$

The reader should note that the new variables in (55) are in the denominator and that their k indices are those of the transformed component $\bar{A}_{\lambda\mu\cdots\omega}$.

7. **Contravariant tensors.** If the n^2 functions A^{ij} of $x_1, \cdots x_n$ on changing the variables to $u_1, \cdots u_n$ go over into

$$\bar{A}^{\lambda\mu} = \sum_{i,j} \frac{\partial u_\lambda}{\partial x_i} \frac{\partial u_\mu}{\partial x_j} A^{ij}, \tag{58}$$

we say they form a *contravariant tensor of order* 2. If we compare (54) with (58), we see the relations between the old and the transformed components differ by having the new variables u in the denominator when covariant and in the numerator when contravariant. The extension of (58) to define contravariant tensors of order k is obvious; instead of 2 partial derivatives we have k.

Example 14. Set $A^1 = dx_1$, $A^2 = dx_2$, $\cdots$ $A^n = dx_n$. The reader should note that 2 in A^2 is an upper index and not an exponent, and so on. On changing the variables these become

$$\bar{A}^1 = du_1, \qquad \bar{A}^2 = du_2, \cdots \bar{A}^n = du_n.$$

But

$$\bar{A}^\lambda = du_\lambda = \sum_i \frac{\partial u_\lambda}{\partial x_i} dx_i = \sum_i \frac{\partial u_\lambda}{\partial x_i} A^i. \tag{59}$$

Hence the n quantities form a contravariant tensor of order 1.

Example 15. By a reasoning unfortunately too long to give here it can be shown that the n^2 quantities a^{ij} defined in (19) form a contravariant tensor of order 2.

8. **Mixed tensors.** Suppose the law of transformation of the n^2 functions of $x_1, \cdots x_n$, call them $A_\beta{}^\alpha$, is defined by

$$\bar{A}_\beta{}^\alpha = \sum_{i,j} \frac{\partial u_\alpha}{\partial x_i}\frac{\partial x_j}{\partial u_\beta} A_j{}^i. \tag{60}$$

If there were l factors with u in the numerator and m factors with u in the denominator, the tensor would be a mixed tensor of order $l + m$, covariant of order m, contravariant of order l.

Example 16. Let us show that the n^4 functions of $x_1, \cdots x_4$

$$G_{\alpha\lambda\mu}{}^\beta = \{\alpha\,\beta,\,\lambda\,\mu\} = \sum_\gamma a^{\beta\gamma} G_{\alpha\gamma\lambda\mu} \tag{61}$$

are the components of a mixed tensor, covariant of order 3 and contravariant of order 1. For, by (57) and (58), we have

$$\begin{aligned}\bar{G}_{\alpha\lambda\mu}{}^\beta &= \sum_\gamma \sum_{i,j} \frac{\partial u_\beta\,\partial u_\gamma}{\partial x_i\;\partial x_j} a^{ij} \sum_{h,k,r,s} \frac{\partial x_h\,\partial x_k\,\partial x_r\,\partial x_s}{\partial u_\alpha\,\partial u_\gamma\,\partial u_\lambda\,\partial u_\mu} G_{hkrs} \\ &= \sum_{i,j,h,k,r,s} (\cdots) \sum_\gamma \frac{\partial u_\gamma\,\partial x_k}{\partial x_j\,\partial u_\gamma}.\end{aligned}$$

From the calculus we know that

$$\sum_\gamma \frac{\partial u_\gamma\,\partial x_k}{\partial x_j\,\partial u_\gamma} = \begin{cases} 1 & \text{if} \quad j = k, \\ 0 & \text{if} \quad j \neq k. \end{cases}$$

Thus the terms in the sum over i, j, k, r, s drop out for which $j \neq k$; therefore

$$\begin{aligned}\bar{G}_{\alpha\lambda\mu}{}^\beta &= \sum_{i,h,r,s} \frac{\partial u_\beta\,\partial x_h\,\partial x_r\,\partial x_s}{\partial x_i\,\partial u_\alpha\,\partial u_\lambda\,\partial u_\mu} \sum_j a^{ij} G_{hjrs} \\ &= \sum_{i,h,r,s} \frac{\partial u_\beta\,\partial x_h\,\partial x_r\,\partial x_s}{\partial x_i\,\partial u_\alpha\,\partial u_\lambda\,\partial u_\mu} G^i{}_{hrs}, \qquad \text{q.e.d.}\end{aligned} \tag{62}$$

We must note one highly important feature which all tensors have in common: The components of the transformed tensor are always *linear* in the components of the original tensor (cf. (53), (55), (57), (58), (59), (60), (61)).

Suppose now a certain law in physics is expressed by the vanishing of the components of a certain tensor, say, for example, by

$$A_{\alpha\beta} = 0. \qquad \alpha, \beta = 1, 2, \cdots n. \tag{63}$$

If we introduce the new variables $u_1, \cdots u_n$, these n^2 equations go over into

$$\bar{A}_{\alpha\beta} = \sum \frac{\partial x_i \, \partial x_j}{\partial u_\alpha \, \partial u_\beta} A_{ij}.$$

But, as each $A_{ij} = 0$ by hypothesis, we have

$$\bar{A}_{\alpha\beta} = 0. \tag{64}$$

Thus the equations (63) hold for any set of coördinates, that is, they are invariant.

9. **Operations on tensors.** The *sum* of two tensors of like character A, B is a tensor whose components are the sum of the components of A and B. Thus the sum of $A = \{A_{\alpha\beta}\}$ and $B = \{B_{\alpha\beta}\}$ has the components $A_{\alpha\beta} + B_{\alpha\beta}$.

The *product of two tensors*, as $A = \{A_{\alpha\beta}\}$ and $B = \{B^\gamma\}$, is the tensor whose components are $C_{\alpha\beta}{}^\gamma = A_{\alpha\beta}B^\gamma$.

The *composition of two tensors* is best illustrated by an example. Suppose $A = \{A_\alpha{}^{\lambda\mu}\}$, $B = \{B^{\alpha\delta\epsilon\eta}_{\lambda\mu\nu}\}$; we set

$$C_\nu{}^{\delta\epsilon\eta} = \sum_{\alpha,\lambda,\mu} A_\alpha{}^{\lambda\mu} B^{\alpha\delta\epsilon\eta}_{\lambda\mu\nu}, \tag{65}$$

the sum extending over the common indices, which must be upper indices in the one tensor and lower in the other. It can be shown easily that the result is a tensor whose character is obtained by cancelling these common indices as indicated in (65).

Example 17. Let $A = \{a_{ij}\}$, $B = \{a^{ir}\}$. Their composition gives

$$\sum_i a_{ij} a^{ir} = a_j{}^r = \begin{cases} 1 & \text{if} \quad r = j, \\ 0 & \text{if} \quad r \neq j, \end{cases} \tag{66}$$

as we saw in (20).

Example 18. The composition of this mixed tensor $\{a_j{}^r\}$ with the mixed tensor $\{G_{\alpha r\mu}{}^j\}$ defined in (61) gives the tensor whose components are

$$\sum_{r,j} a_j{}^r G_{\alpha r\mu}{}^j = \sum_r G_{\alpha r\mu}{}^r = \sum_r \{\alpha\, r,\, r\, \mu\} = G_{\alpha\mu}. \tag{67}$$

This tensor is historic. In fact the equations

$$G_{\alpha\mu} = 0, \qquad \alpha, \mu = 1, 2, 3, 4, \tag{68}$$

determine the metric (9) of the space about the sun.

Contraction. This is another operation which leads to a tensor. Suppose, for example, in a mixed tensor whose components are $A_{\lambda\mu\nu}{}^{\alpha\beta}$; we set $\alpha = \lambda$, $\beta = \nu$ and sum over α, β, thus

$$\sum_{\alpha,\beta} A_{\alpha\mu\beta}{}^{\alpha\beta} = A_\mu. \tag{69}$$

This is found to be a tensor whose character is obtained by dropping the common upper and lower indices. Thus (69) are the components of a covariant tensor of order 1 as indicated.

Example 19. Contracting $\{G_{\alpha r\mu}{}^{j}\}$ by setting $j = r$ we get a tensor whose components are

$$\sum_r G_{\alpha r\mu}{}^{r} = \sum_r \{\alpha\, r,\, r\, \mu\} = G_{\alpha\mu},$$

the same tensor obtained by composition in example 18.

10. **Tensors of order 0.** If we compound $\{A_{ij}\}$ with $\{B^{ij}\}$, we get a tensor whose sole component is

(70) $$\sum_{i,j} A_{ij} B^{ij}.$$

On transforming this becomes

(71) $$\begin{aligned}\sum_{i,j} \overline{A}_{ij}\overline{B}^{ij} &= \sum_{i,j}\sum_{\alpha,\beta} \frac{\partial x_\alpha}{\partial u_i}\frac{\partial u_\beta}{\partial x_j} A_{\alpha\beta} \sum_{\lambda,\mu} \frac{\partial u_i}{\partial x_\lambda}\frac{\partial u_j}{\partial x_\mu} A^{\lambda\mu} \\ &= \sum_{\alpha,\beta,\lambda,\mu} A_{\alpha\beta} A^{\lambda\mu} \sum_i \frac{\partial x_\alpha}{\partial u_i}\frac{\partial u_i}{\partial x_\lambda} \sum_j \frac{\partial x_\beta}{\partial u_j}\frac{\partial u_j}{\partial x_\mu}.\end{aligned}$$

Now $\sum_j = 0$ or 1 according as $\mu = \beta$ or does not, and a similar remark holds for the sum $\sum_i$. Thus (71) reduces to

$$\sum \overline{A}_{ij}\overline{A}^{ij} = \sum A_{ij} A^{ij},$$

i.e., the expression is an invariant. On the other hand (70) is a tensor whose character is obtained by omitting common upper and lower indices; as no index is left, we may regard it as a tensor of order 0. The foregoing may be extended obviously to tensors of any order.

Example 20.

$$ds^2 = \sum a_{ij}\, dx_i dx_j.$$

Here a_{ij} is covariant of order 2, $\{dx_i dx_j\} = \{B^{ij}\}$ is a contravariant tensor of order 2 also. As ds^2 is obtained by compounding these two tensors, it is an invariant.

Example 21.

$$\cos\theta = \sum_{i,j} a_{ij} \frac{dx_i}{ds}\frac{\delta x_j}{\delta s}.$$

This is the composition of $\{a_{ij}\}$ with $\left\{\dfrac{dx_i}{ds}\dfrac{\delta x_j}{\delta s}\right\} = \{B^{ij}\}$. Hence $\cos\theta$ is an invariant, as already observed.

Example 22. *Beltrami's parameter.*

(72) $$\Delta_1(\varphi) = \sum_{i,j} a^{ij} \frac{\partial\varphi}{\partial x_i}\frac{\partial\varphi}{\partial x_j}.$$

Let us first show that $A_i = \partial\varphi/\partial x_i$ are the components of a covariant tensor of order 1. In fact

$$A_i = \frac{\partial\varphi}{\partial u_i} = \frac{\partial\varphi}{\partial x_1}\frac{\partial x_1}{\partial u_i} + \cdots + \frac{\partial\varphi}{\partial x_n}\frac{\partial x_n}{\partial u_i}; \qquad \therefore \overline{A}_i = \sum_\alpha \frac{\partial x_\alpha}{\partial u_i} A_i.$$

Thus $B_{ij} = \dfrac{\partial\varphi}{\partial x_i}\dfrac{\partial\varphi}{\partial x_j}$ are the components of a covariant tensor of order 2. Hence (72) is an invariant.

Example 23. *Beltrami's mixed parameter.*

$$\nabla(\varphi, \psi) = \sum_{ij} a^{ij}\frac{\partial\varphi}{\partial x_i}\frac{\partial\psi}{\partial x_j} \tag{73}$$

is obviously an invariant.

Example 24.

$$\sum_{r,s} a_{rs}a^{rs}.$$

This is also an invariant. In fact by (66)

$$\sum_{r,s} a_{rs}a^{rs} = \sum_r \sum_s a_{rs}a^{rs} = \sum_r a_r{}^r = n. \tag{74}$$

Example 25.

$$G = \sum_{\lambda,\mu} a^{\lambda\mu}G_{\lambda\mu}, \tag{75}$$

where by (67) $G_{\lambda,\mu} = \sum_h \{\lambda h, h \mu\}$ is an invariant. This invariant is fundamental in Einstein's theory, as we shall see. It is called the *curvature invariant.* For a euclidean space G is zero.

11. **Covariant differentiation.** Let $\{A_i\}$ be a tensor of order 1. We find that

$$A_{\alpha\lambda} = \frac{\partial A_\alpha}{\partial x_\lambda} - \sum_h \begin{Bmatrix} \alpha\ \lambda \\ h \end{Bmatrix} A_h \tag{76}$$

are the components of a covariant tensor of order 2. Similarly

$$A_\lambda{}^\alpha = \frac{\partial A^\alpha}{\partial x_\lambda} + \sum_h \begin{Bmatrix} h\ \lambda \\ \alpha \end{Bmatrix} A^h \tag{77}$$

are the components of a contravariant tensor of order 2. These tensors we say are obtained from $\{A_\alpha\}$ and $\{A^\alpha\}$ by *covariant differentiation.* It is easy to extend this process to tensors of any order. Thus the covariant derivative of the three types of tensors of order 2 relative to x_λ are the tensors of order 3 whose components are

$$A_{\alpha\beta\lambda} = \frac{\partial A_{\alpha\beta}}{\partial x_\lambda} - \sum_h \begin{Bmatrix} \alpha\ \lambda \\ h \end{Bmatrix} A_{h\beta} - \sum_h \begin{Bmatrix} \beta\ \lambda \\ h \end{Bmatrix} A_{\alpha h}, \tag{78}$$

$$A_\lambda{}^{\alpha\beta} = \frac{\partial A^{\alpha\beta}}{\partial x_\lambda} + \sum_h \begin{Bmatrix} h\ \lambda \\ \alpha \end{Bmatrix} A^{h\beta} + \sum_h \begin{Bmatrix} h\ \lambda \\ \beta \end{Bmatrix} A^{\alpha h}, \tag{79}$$

$$A^\alpha_{\beta\lambda} = \frac{\partial A_\beta{}^\alpha}{\partial x_\lambda} - \sum_h \begin{Bmatrix} \beta\ \lambda \\ h \end{Bmatrix} A_h{}^\alpha + \sum_h \begin{Bmatrix} h\ \lambda \\ \alpha \end{Bmatrix} A^h_\beta. \tag{80}$$

It is important to note that the covariant derivatives of the fundamental tensor $\{a_{ij}\}$ are all zero. Let us note also that when these com-

ponents a_{ij} are constants, covariant differentiation is identical with ordinary differentiation, since the Christoffel symbols $\{{}^{\alpha\beta}_{\gamma}\}$ are all zero.

Example 26. Let F be a function of $x_1, \cdots x_n$ and set $F_i = \partial F/\partial x_i$. Then by (76)

$$F_{ik} = \frac{\partial^2 F}{\partial x_i \partial x_k} - \sum_h \left\{ \begin{matrix} i\ k \\ h \end{matrix} \right\} \frac{\partial F}{\partial x_h}.$$

If we compound this tensor with the tensor whose components are a^{ik}, we get *Beltrami's second differential parameter*, which is therefore an invariant; viz.,

$$\Delta_2 F = \sum_{i,k} a^{ik} F_{ik} = \sum_{i,k} a^{ik} \left[\frac{\partial^2 F}{\partial x_i \partial x_k} - \sum_h \left\{ \begin{matrix} i\ k \\ h \end{matrix} \right\} \frac{\partial F}{\partial x_h} \right]. \tag{81}$$

When $ds^2 = dx_1^2 + dx_2^2 + dx_3^2$, this reduces to

$$\Delta_2 F = \frac{\partial^2 F}{\partial x_1^2} + \frac{\partial^2 F}{\partial x_2^2} + \frac{\partial^2 F}{\partial x_3^3}.$$

12. **Divergence.** In the restricted theory of relativity the divergence of certain tensors is of fundamental importance. They are equally important in Einstein's theory. Consider, for example, the covariant tensor whose components are $A_{\lambda\mu}$. Its covariant derivative relative to x_k has the components $A_{\lambda\mu\,k}$. It is of order 3. To get a tensor of order 1 we compound it with the fundamental tensor $\{a^{\mu k}\}$ getting a covariant tensor of order 1 whose components are

$$\sum_{k,\mu} a^{\mu k} A_{\lambda\mu\,k} = A_\lambda; \tag{82}$$

we call this tensor the *divergence* of $\{A_{\lambda\mu}\}$ and write it div $\{A_{\lambda\mu}\}$. Similarly the divergence of the contravariant tensor $\{A^{\lambda\mu}\}$ relative to x_k has the components

$$\sum_{k,\mu} a_\mu{}^k A_k{}^{\lambda\mu} = \sum_k A_k{}^{\lambda k} = A^\lambda, \tag{83}$$

obviously a contravariant tensor of order 1. In a similar manner we may define the divergence of any tensor, but, as we shall not need them, we will not take space to write them down.

13. **Einstein's metric.** We have seen that the metric of Einstein's 4-way space is determined by a quadratic differential form

$$ds^2 = \sum a_{ij} dx_i dx_j, \qquad i, j = 1, 2, 3, 4, \quad a_{ij} = a_{ji}. \tag{84}$$

As yet, however, the 10 coefficients a_{ij} are undetermined functions of the 3 space coördinates x_1, x_2, x_3 and the time coördinate x_4. To determine these a's Einstein makes use of the fact that the restricted relativity theory gives a very satisfactory account of a wide class of phenomena. In this theory the metric is given by

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2. \tag{85}$$

Einstein, therefore, requires as a first restriction that (84) shall reduce to (85) by a suitable transformation of the variables and for a sufficiently small region about a given point, i.e., neglecting infinitesimals of a higher order. This amounts to Einstein's celebrated *principle of equivalence*. The further determination of the a's depends upon the presence of material bodies and electricity. For brevity we shall consider only a special case of a gravitational field. In a system of bodies removed from all other influences, i.e., a *complete system*, the most important facts relate to the conservation of energy and momentum. In the restricted theory this is expressed by the vanishing of the energy-momentum tensor T of that theory. Einstein carries this over and takes, as a generalization of T, a symmetric tensor covering a wide class of phenomena, whose components are in contravariant form

$$T^{\lambda\mu} = \rho \frac{dx_\lambda}{ds}\frac{dx_\mu}{ds}, \tag{86}$$

or in the equivalent covariant form

$$T_{ij} = \sum_{\lambda,\mu} \rho a_{i\lambda} a_{j\mu} \frac{dx_\lambda}{ds}\frac{dx_\mu}{ds}, \tag{86a}$$

where ρ is the density of matter and ds is given by (84). Thus Einstein requires

$$\operatorname{div}\{T^{ij}\} = 0 \qquad \text{or} \qquad \operatorname{div}\{T_{ij}\} = 0, \tag{87}$$

either one of these equations having the other as a consequence.*

* A few words of explanation may be welcome to some readers. Let us recall the equations of motion of an elastic body (e.g., viscous fluid). At the point $x = (x_1, x_2, x_3)$ let the components of the stress p_i on a plane perpendicular to the x_i axis be denoted by p_{ij}, $j = 1, 2, 3$. Let u_i be the components of the velocity of the element of mass dm of density ρ at the point x. Then, when external forces are neglected,

$$\sum_j \frac{\partial p_{ij}}{\partial x_j} + \rho \frac{du_i}{dt} = 0, \qquad i = 1, 2, 3. \tag{a}$$

Here $du_i/dt = \partial u_i/\partial t + \Sigma_j u_j \partial u_i/\partial x_j = Du/Dt$ in English works. To these we add the equation of continuity

$$\frac{\partial \rho}{\partial t} + \sum_j \frac{\partial(\rho u_j)}{\partial x_j} = 0. \tag{b}$$

The four equations (a), (b) determine the four unknowns ρ, u_1, u_2, u_3.

Let us see how these equations look in the restricted theory of relativity. For simplicity we shall choose our units so that the velocity of light in vacuo $c = 1$. If we set

$$q_{ij} = p_{ij} + \rho u_i u_j, \qquad i, j = 1, 2, 3, \tag{c}$$

the energy-momentum tensor T has the components

$$\begin{matrix} q_{11} & q_{12} & q_{13} & \rho u_1, \\ q_{21} & q_{22} & q_{23} & \rho u_2, \\ q_{31} & q_{32} & q_{33} & \rho u_3, \\ \rho u_1 & \rho u_2 & \rho u_3 & \rho. \end{matrix} \tag{d}$$

This is a *physical* requirement. The question now is, how does this gravitating matter affect the metric of the surrounding space? In the older mechanics the gravitational field is determined by Poisson's equation,

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -4\pi\rho. \tag{88}$$

The left side of (88) is a linear function of the second derivatives of the potential function V, and this function, as the right side of (88) shows, is proportional to the amount of matter per unit volume. Now, in the restricted theory of relativity, mass and energy are proportional. This leads one to generalize by assuming that the effect of matter on the metric of space is obtained by setting the energy momentum tensor proportional to a space tensor of order 2 and, by analogy to (88), we shall take one which is linear in the second derivatives of a_{ij}. The most natural tensor of this kind to take would be $\{G_{ij}\}$ defined in (67), but, unfortunately, the divergence of this tensor is not 0 and hence a relation of the type $G_{ij} = \alpha T_{ij}$ would contradict (87). From G_{ij} we can however deduce a tensor whose divergence is zero by adding the term $-\frac{1}{2}a_{ij}G$, where G

The motion of the body is now determined by the equation

$$\text{div } T = 0, \tag{e}$$

where, in general, the divergence of a tensor A whose components are $A^{\lambda\mu}$, $\lambda, \mu = 1, 2, 3, 4$ is a vector whose components are

$$\frac{\partial A^{\lambda 1}}{\partial x_1} + \frac{\partial A^{\lambda 2}}{\partial x_2} + \frac{\partial A^{\lambda 3}}{\partial x_3} + \frac{\partial A^{\lambda 4}}{\partial x_4}. \tag{f}$$

Thus the equation (e) is equivalent to four equations. For $\lambda = i = 1, 2, 3$, $x_4 = t$ it gives $\sum_j \partial T^{ij}/\partial x_j + \partial(\rho u_i)/\partial t$ or, using (c),

$$\sum_j \frac{\partial p_{ij}}{\partial x_j} + \sum_j \frac{\partial(\rho u_i u_j)}{\partial x_j} + \frac{\partial(\rho u_i)}{\partial t} = 0. \tag{g}$$

For $\lambda = 4$, (e) gives

$$\sum_j \frac{\partial(\rho u_j)}{\partial x_j} + \frac{\partial \rho}{\partial t} = 0. \tag{h}$$

We note that equation (h) has the same form as (b). To reduce (g) to the form (a) we observe

$$\sum_j \frac{\partial(\rho u_i u_j)}{\partial x_j} + \frac{\partial(\rho u_i)}{\partial t} = u_i \sum \frac{\partial(\rho u_j)}{\partial x_j} + \rho \sum u_j \frac{\partial u_i}{\partial x_j} + \rho \frac{\partial u_i}{\partial t} + u_i \frac{\partial \rho}{\partial t}$$

or, using (h),

$$= \rho \left\{ \frac{\partial u_i}{\partial t} + \sum_j u_j \frac{\partial u_i}{\partial x_j} \right\} = \rho \frac{du_i}{dt}.$$

Hence $\sum_j \partial p_{ij}/\partial x_j + \rho du_i/dt = 0$, which has the same form as (a).

To show that (e) is a special case of Einstein's equation (87) or div $\{T^{ij}\} = 0$, we recall that, when the coefficients a_{ij} in (84) are constants, as they are in the restricted theory of relativity, covariant differentiation and ordinary differentiation are the same. Then by (83) the components of div $\{T^{ij}\}$ are

$$\frac{\partial T^{\lambda 1}}{\partial x_1} + \frac{\partial T^{\lambda 2}}{\partial x_2} + \frac{\partial T^{\lambda 3}}{\partial x_3} + \frac{\partial T^{\lambda 4}}{\partial x_4}, \qquad \lambda = 1, 2, 3, 4,$$

which are identical with (f).

is the curvature tensor defined by (75). The required relation is thus obtained by setting the space tensor $G_{ij} - \frac{1}{2}a_{ij}G$ proportional to the energy momentum tensor T_{ij}. We have therefore as the 10 equations to determine the 10 unknown a_{ij}

(89) $$G_{ij} - \tfrac{1}{2}a_{ij}G = -\kappa T_{ij}, \qquad i, j = 1, 2, 3, 4.$$

We can give these equations another form which is useful. Multiplying (89) by a^{ij} and summing we get

(90) $$\sum_{i,j} a^{ij}G_{ij} - \tfrac{1}{2}G\sum a_{ij}a^{ij} = -\kappa\sum a^{ij}T_{ij}.$$

Now by (74) $\sum_{i,j} a_{ij}a^{ij} = 4$, since $n = 4$. Hence, if we set

$$T = \sum_{i,j} a^{ij}T_{ij},$$

the energy momentum invariant, we get $G - 2G = T$, which, set in (90), gives

(91) $$G = \kappa T.$$

Putting this in (89) gives the desired relation

(92) $$G_{ij} = -\kappa(T_{ij} - \tfrac{1}{2}a_{ij}T).$$

When there is no matter present at the point x, T and T_{ij} vanish. Thus for space outside gravitating matter the 10 coefficients a_{ij} are determined by the 10 differential equations

(93) $$G_{ij} = 0,$$

which are linear in the second partial derivatives of the $a_{\lambda\mu}$. By means of these equations together with the radial symmetry of space we may show that the metric of our space produced by the gravitation of a central body, as the sun, may be given the form expressed in (9). The constant κ which figures in (89) and (92) is found to have the value

(94) $$\kappa = 8\pi\frac{k^2}{c^2} = 2\cdot 10^{-27} \text{ c.g.s. units.}$$

14. **Cosmological considerations.** In studying the behavior of a complete system it is often a great convenience in ordinary mathematical physics to replace the boundary conditions by giving their values at infinity. This device was used by Einstein in his celebrated paper on the perihelion of Mercury (1915). He supposed the a_{ij} to take on the values at infinity (for a proper set of coördinates) given by the scheme

(95) $$\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{matrix}$$

which correspond to the metric $ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$ of the restricted theory of relativity. A disadvantage of this assumption lies in the fact that these values are tied down to a certain set of coördinates, they are not invariant. For this and other reasons (stability of our stellar system) Einstein was led to adopt a universe of finite magnitude, i.e., an elliptic metric. In this space there are no boundary values as there is no boundary. He supposes that matter is on the average uniformly distributed of density ρ. The stars are concentrations of this matter, whose greater density is compensated by a rarity of matter elsewhere. As the metric of his 4-way space he takes

$$(96) \qquad ds^2 = c^2dt^2 - \frac{dx_1^2 + dx_2^2 + dx_3^2}{\left[1 + \frac{k^2}{4}(x_1^2 + x_2^2 + x_3^2)\right]^2}.$$

It turns out, however, that this metric is in conflict with the equations (89) if we assume that the world matter has a velocity small in comparison with light, an assumption justified by the relatively small velocity of the stars so far as ascertained. This difficulty is easily remedied by introducing a new term in (89). For the left-hand side of (89) was chosen as the simplest covariant tensor of order 2 whose divergence was zero. Now we saw that the covariant derivatives of the fundamental tensor, viz., $a_{ij/k}$, are all zero. Hence by (82) div $a_{ij} = 0$. Thus we may add a term λa_{ij} (λ = constant) to the left side of (89) and still have a tensor whose divergence vanishes. Einstein therefore sets

$$(97) \qquad G_{ij} - \lambda a_{ij} - \tfrac{1}{2}a_{ij}G = -\kappa T_{ij},$$

which is now in harmony with (96). As before we now find

$$(98) \qquad G = 4\lambda + \kappa T,$$

which, set in (97), gives

$$(99) \qquad G_{ij} - \lambda a_{ij} = -\kappa(T_{ij} - \tfrac{1}{2}a_{ij}T).$$

Where there is no matter, T and T_{ij} vanish and (99) becomes

$$(100) \qquad G_{ij} - \lambda a_{ij} = 0,$$

which takes the place of the former equations (93).

The two universal constants κ, λ are related by

$$(101) \qquad \lambda = \tfrac{1}{2}\kappa\rho, \qquad \kappa\rho = \frac{2}{R^2},$$

where $1/R = k$ is the curvature of the x_1, x_2, x_3 space R_3 obtained by setting t = const. in (96).

15. **Estimation of the size of the universe.** Astronomers often use as a unit of distance 1 parsec which equals the distance of a star whose parallax

is 1″. Thus 1 parsec $= 2 \cdot 10^5$ orbrads $= 2 \cdot 10^5 \cdot 150 \cdot 10^6$ kilometers, or 1 parsec $= 3 \cdot 10^{15}$ cm. Let us now assume with Kapteyn that the density of the cosmos is about the same as in a cube described about the sun and having a side of 10 parsecs $= 3 \cdot 10^{19}$ cm. The volume of the cube is therefore $27 \cdot 10^{57}$ cm.^{3}. In such a cube Kapteyn estimates that there are about 80 suns of about the mass of ours. As the mass of our sun is about $2 \cdot 10^{33}$ gm., the mass of these suns is $16 \cdot 10^{34}$ gms., we have

$$\text{density} = \rho = \frac{\text{mass}}{\text{volume}} = \frac{16 \cdot 10^{34}}{27 \cdot 10^{57}} = 5,\, 9 \cdot 10^{-24}.$$

From (101) we have $R^2 = 2\, \kappa\rho$ and, as by (94) $\kappa = 2 \cdot 10^{-7}$, we have

$$R^2 = \frac{2}{2 \cdot 10^{-27}} \cdot \frac{1}{5,\, 9 \cdot 10^{-24}} = 1,\, 7 \cdot 10^{50},$$

therefore

$$R = 1,\, 3 \cdot 10^{25} \text{ cm.} = 9 \cdot 10^{11} \text{ orbrads.}$$

CAUCHY'S PAPER OF 1814 ON DEFINITE INTEGRALS.*

BY H. J. ETTLINGER.

Introduction. In 1814 Augustin Louis Cauchy presented before the Académie des Sciences a "memoir on definite integrals," in which appears for the first time the essence of his discoveries on residues. The memoir was first printed in 1825† with additional notes and again in 1882‡ with no change save in the matter of notation.

Although the kernel of the idea, that the integral of an analytic function of a complex variable taken along a closed path depends entirely upon the behavior of the function at points of discontinuity within the path, is contained in the paper, yet there are several reasons why the reader might notice nothing at all like this theorem. In the first place, although geometrical representation is now an essential feature of every presentation of the theory of functions, Cauchy used neither figures nor geometrical language. In the second place, the fundamental theorem, and, indeed, all the applications in this paper, concern simple integrals; but the author states the central problem as the determination of the difference in the value of an iterated integral according to the order of integration with respect to the two variables. By the use of this difference he obtains the residue, thereby obscuring the relation of the latter to a line integral. Thirdly, he refrains from using complex quantities, invariably separating an equation into its real and imaginary parts. This necessitates longer equations, more of them, clumsier notation, and a much more obscure treatment than would be the case had he used complex quantities. Cauchy himself came to appreciate this fact, for his footnotes of 1825 are devoted to the simpler complex equations from which his real ones can be readily deduced. Finally, all editions abound in misprints.

For these reasons the discoveries contained in this memoir were not appreciated even by the great mathematicians of his time. Poisson§ saw in the paper merely a means of evaluating integrals and remarked that, at least so far as the first part was concerned, no new formulæ were announced. As for the evaluation of iterated integrals by the so-called

* Presented to the Amer. Math. Soc., Sept. 2, 1919.

† "Mémoire sur les intégrales définies," Savants Étrangers, 1, p. 509, Académie des Sciences de l'Institut de France.

‡ Œuvres Complétes, I série, 1, p. 319 ff.

§ Bulletin de la Societé Philomatique (3), 1, 1814, p. 185.

"singular" integrals (which are equal to the difference between the value obtained by integrating first with respect to x and then with respect to y and the value obtained by integrating in the reverse order) he said that, though the new method was worthy of consideration, it ought not to replace the *old ones*! Lacroix and Legendre, in the official report on the paper, stated as the valuable results obtained by Cauchy: (1) the construction of a series of general formulæ for transforming and evaluating definite integrals, (2) the pointing out of the fact that the value of an iterated integral may depend on the order of integration, (3) the discovery of the cause and amount of this difference in value, (4) the derivation of new formulæ which, to be sure, might have been otherwise obtained. It seems, then, likely that the foremost mathematicians of that time failed to recognize the contributions of main importance in this paper.

To appreciate thoroughly the memoir, the following facts must be noted in addition: (1) imaginaries had no secure arithmetical basis in 1814, (2) this was the first deduction by rigorous methods of the formulæ, hitherto obtained by purely formal processes, for evaluating definite integrals, (3) while the form had not yet been cast in the ϵ-mould, which itself originated with Cauchy, nevertheless the proofs are so conceived that they correspond in substance to the standards of rigor of the present day.

Part I.

Continuous integrand. In discussing the memoir we shall frequently combine two separate real equations into one complex equation, as Cauchy did in his notes of 1825 and, very likely, in his original work. We shall also adopt the language of modern analysis for the sake of clearness and accuracy.

The first theorem proved in the memoir is, in effect, that if a function of a complex variable is analytic throughout a region of a certain type and continuous in and on the boundary, the integral of the function taken along the boundary of the region is zero.* The regions considered are mapped in a one-to-one manner and continuously, but not in general conformally, on a rectangle in the real (x, y) plane. The mapping on the complex $M + Ni$ plane is performed by taking M and N as real and continuous functions of x and y with derivatives of all orders with respect to x and y, continuous in x and y regarded as independent variables.

* For modern treatment of this theorem see Osgood, Lehrbuch der Funktionentheorie, erster Band, zweite Auflage, pp. 284–285; Pierpont, Functions of a Complex Variable, pp. 211–214; Goursat, Cours d'Analyse Mathématique, tome II, pp. 82–92. These three text-books will hereafter be referred to as O., P., G., respectively.

Let*

(1) $$f(M + Ni) = P + Qi$$

be an analytic function of $M + Ni$ in a certain region, S, of the $M + Ni$ plane, and let

$$M = \phi(x, y) \qquad \text{and} \qquad N = \psi(x, y)$$

be single-valued functions, continuous in x and y, in a rectangle, R ($0 \leqq x \leqq a$, $0 \leqq y \leqq b$), and on the boundary, Γ, and possessing continuous partial derivatives of all orders with respect to x and y in R and on Γ.

Furthermore, let†

(2) $$S + Ti = f(M + Ni)\frac{\partial(M + Ni)}{\partial x},$$

(3) $$U + Vi = f(M + Ni)\frac{\partial(M + Ni)}{\partial y}.$$

Differentiate (2) and (3) with regard to y and x respectively:

$$\frac{\partial S}{\partial y} + i\frac{\partial T}{\partial y} = f'(M + Ni)\frac{\partial(M + Ni)}{\partial x}\cdot\frac{\partial(M + Ni)}{\partial y} + f(M + Ni)\frac{\partial^2(M + Ni)}{\partial y\partial x},$$

$$\frac{\partial U}{\partial x} + i\frac{\partial V}{\partial x} = f'(M + Ni)\frac{\partial(M + Ni)}{\partial y}\cdot\frac{\partial(M + Ni)}{\partial x} + f(M + Ni)\frac{\partial^2(M + Ni)}{\partial x\partial y}.$$

But under the conditions imposed

$$\frac{\partial^2(M + Ni)}{\partial x\partial y} = \frac{\partial^2(M + Ni)}{\partial y\partial x}\ddagger$$

or

$$\frac{\partial S}{\partial y} + \frac{i\partial T}{\partial y} = \frac{\partial U}{\partial x} + \frac{i\partial V}{\partial x}.$$

Hence

(4) $$\frac{\partial S}{\partial y} = \frac{\partial U}{\partial x} \qquad \text{and} \qquad \frac{\partial T}{\partial y} = \frac{\partial V}{\partial x}.$$

Multiplying the equations (4) by $dydx$ and integrating from $x = 0$ to $x = a$ and $y = 0$ to $y = b$, and noting further that, since the integrand is continuous, the order of integration can be reversed, we have:

* The notation of the original paper has been changed here from $P' + P''i$ to $P + Qi$ and $y = x + zi$ to $z = x + yi$.

† Hereafter $\frac{\partial(M + Ni)}{\partial x} - i\frac{\partial(M + Ni)}{\partial y}$ will be designated by $\frac{d(M + Ni)}{d(x + yi)}$.

‡ See Goursat-Hedrick, Mathematical Analysis, vol. I, p. 13.

$$(5) \qquad \int_0^a dx \int_0^b \frac{\partial S}{\partial y} dy = \int_0^b dy \int_0^a \frac{\partial U}{\partial x} dx.$$

Let $S(x, b) = S$, $S(x, 0) = s$, $U(a, y) = U$, $U(0, y) = u$; then equation (5) becomes

$$(6) \qquad \int_0^a Sdx - \int_0^a sdx = \int_0^b Udy - \int_0^b udy.$$

In a similar manner, letting $T(x, b) = T$, $T(x, 0) = t$ and $V(a, y) = V$, $V(0, y) = v$, we obtain

$$(7) \qquad \int_0^a Tdx - \int_0^a tdx = \int_0^b Vdy - \int_0^b vdy.$$

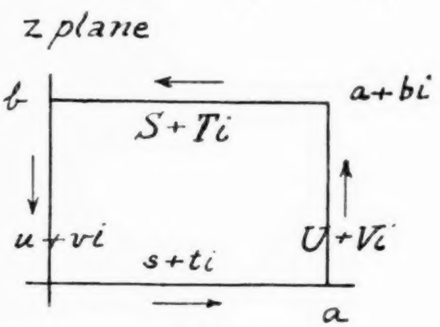


Fig. 1.

Multiplying (7) by $-i$ and (6) by -1 and adding we have

$$\int_0^a (s + ti)dx + \int_0^b (U + Vi)dy - \int_0^a (S + Ti)dx - \int_0^b (u + vi)dy = 0,$$

or

$$\int_L f(M + Ni) \frac{d(M + Ni)}{d(x + yi)} d(x + yi) = 0,$$

which means that around the rectangle here given in the z plane, and hence in the $M + Ni$ plane about the corresponding curve,* L_1,

$$(8) \qquad \int_L f(M + Ni) d(M + Ni) = 0.$$

Hence

Fundamental Theorem I: *Let $f(M + Ni)$ be an analytic function of $M + Ni$ in a certain region, S, of the $M + Ni$ plane, continuous in S, and on the boundary, L, and let $M = \phi(x, y)$ and $N = \psi(x, y)$ be single-valued continuous functions of x and y in a rectangle, R $(0 \leqq x \leqq a,\ 0 \leqq y \leqq b)$, and on the boundary, Γ, possessing continuous partial derivatives of*

* Cf. equation (3) and equations (A), footnote, Œuvres Complétes, I série, 1, p. 338. This memoir will be referred to hereafter as O.C.

all orders with respect to x and y and mapping the closed region, S, on the closed rectangle, R, in a one-to-one manner and continuously; then the integral of $f(M + Ni)d(M + Ni)$ taken around L in the positive direction vanishes, or

$$\int_L f(M + Ni)d(M + Ni) = 0.$$

In the applications Cauchy uses the real equations in S and U, T and V respectively and does not combine them as is here done. The functions used in this paper for M and N and the corresponding maps of the rectangle, R, on the $M + Ni$ plane are given in figures 2, 3, 4, and 5.

$$1°. \quad M = x, \qquad N = y.$$

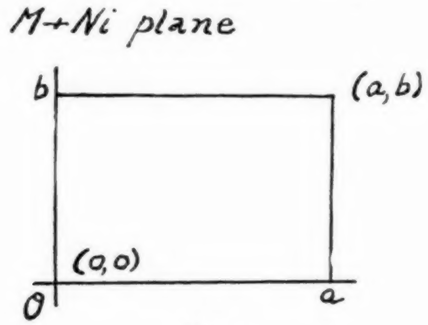


FIG. 2.

$$2°. \quad M = ax, \qquad N = xy;$$
$$a > 0.$$

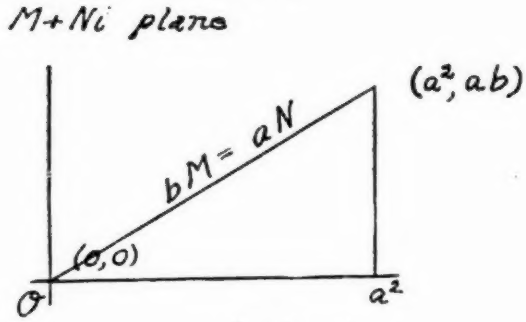


FIG. 3.

In several of the applications Cauchy allows a, the upper limit of the x-interval, to become infinite. He considered that his conclusions could be extended to this case if the function $f(M + Ni)$ approaches a limit for each value of y $(0 \leqq y \leqq b)$ when a becomes infinite and the improper integrals thus introduced converge. These conditions are, of course,

3°. $M = x \cos y, \qquad N = x \sin y.$

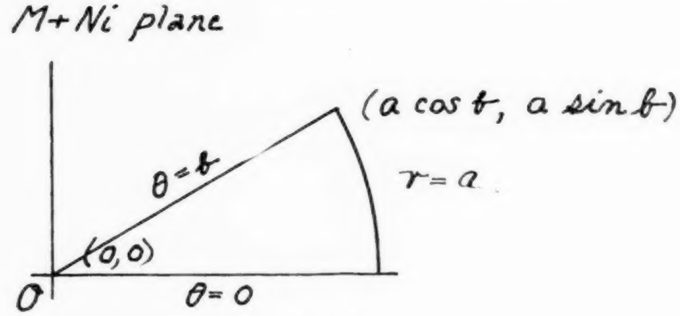


FIG. 4.

4°. $M = ax^2, \qquad N = xy;$
$a > 0.$

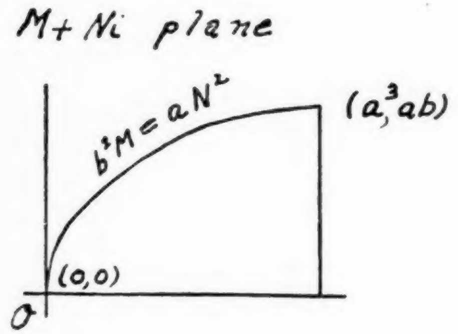


FIG. 5.

insufficient, but since all the functions $f(M + Ni)$ considered by Cauchy approach their limits uniformly in $(0 \leqq y \leqq b)$ and the integrals converge, the results may be established as correct.

The following is an example of the method of application and of the results of Part I.

Region 1°. (See Fig. 2.)

$$M = x, \qquad N = y.$$

Let $f(x + yi) = P(x, y) + Q(x, y)i$, such that $Q(x, 0) \equiv 0$, and $S + Ti = P + Qi$, $U + Vi = -Q + Pi$.

Equations (6) and (7) yield

$$(6') \quad \int_0^a P(x, b)dx - \int_0^a P(x, 0)dx + \int_0^b Q(a, y)dy - \int_0^b Q(0, y)dy = 0,$$

$$(7') \quad \int_0^a Q(x, b)dx - \int_0^b P(a, y)dy + \int_0^b P(0, y)dy = 0.$$

Apply these equations to

$$f(z) = e^{-z^2} \qquad \text{where} \qquad z = x + yi.^*$$
$$P(x, y) = e^{-x^2}e^{y^2} \cos 2xy, \qquad Q(x, y) = e^{-x^2}e^{y^2} \sin 2xy,$$
$$P(x, 0) = e^{-x^2}, \qquad P(0, y) = e^{y^2}, \qquad Q(0, y) = 0,$$

so that in this case equations (6′) and (7′) become respectively

$$(9) \qquad \int_0^a e^{-x^2}e^{b^2} \cos 2bx dx - e^{-a^2}\int_0^b e^{y^2} \sin 2ay dy = \int_0^a e^{-x^2} dx,$$

$$(10) \qquad -\int_0^a e^{-x^2}e^{b^2} \sin 2bx dx - e^{-a^2}\int_0^b e^{y^2} \cos 2ay dy = -\int_0^b e^{y^2} dy.$$

Now let a increase without limit. The second integral in each of the equations (9) and (10) vanishes, for

$$\left|\int_0^b e^{-a^2}e^{y^2} \sin 2ay dy\right| \leqq \int_0^b \left|\sin 2ay\right| e^{-a^2}e^{y^2} dy$$
$$\leqq e^{-a^2}\int_0^b e^{y^2} dy \leqq b e^{b^2-a^2}, \qquad b > 0.$$

But

$$\lim_{a \to \infty} b e^{b^2-a^2} = 0,$$

hence

$$\lim_{a \to \infty} \int_0^b e^{-x^2}e^{y^2} \sin 2ay dy = 0.$$

Similarly

$$\lim_{a \to \infty} \int_0^b e^{-a^2}e^{y^2} \cos 2ay dy = 0.$$

Since it can be readily shown that the other integrals converge, we are justified in writing Cauchy's equations:

$$\int_0^\infty e^{-x^2} \cos 2bx dx = e^{-b^2}\int_0^\infty e^{-x^2} dx = \frac{e^{-b^2}\sqrt{\pi}}{2},$$

$$\int_0^\infty e^{-x^2} \sin 2bx dx = e^{-b^2}\int_0^b e^{y^2} dy,$$

if we assume $\int_0^\infty e^{-x^2} dx = \sqrt{\pi}/2$.

PART II.

Integrands with poles. In the second part of the memoir Cauchy deals with the integrals of functions which are discontinuous at isolated points. In all the applications these singularities are simple poles.†

* Cf. O., p. 293, Beispiel, 4. G., p. 121, 3°.
† O.C., p. 413.

Here Cauchy obtains for the first time a formula which contains the essence of the theorem on residues. The true significance of Cauchy's method at this point is very obscure. The result is apparently stated in terms of the evaluation in two different orders of an iterated integral whose integrand has a singularity at a single point.* As a matter of fact, the iterated integral plays an unessential rôle in Part II, since all the theorems and applications are concerned with simple integrals only. Moreover no useful facts are developed concerning iterated integrals.

The exposition and criticism of Cauchy's method we lay aside for the moment and proceed to set forth a method by which the results of Part II are very simply obtained from the fundamental theorem of Part I. This method is not so very unlike Cauchy's, as will be pointed out later, and would probably be used by him were he writing in the notation of the present-day analyst. It is the method used in many modern text books on the Theory of Functions.†

FUNDAMENTAL THEOREM II: *Let $f(M + Ni)$ be an analytic function of $M + Ni$ in a certain region, S, of the $M + Ni$ plane, except for a single pole at $m + ni$, inside of S, and continuous in and on the boundary, L, of S, except at this pole. Let $M = \phi(x, y)$ and $N = \psi(x, y)$ be continuous functions of x and y in and on the boundary, Γ, of a rectangle, R $(0 \leqq x \leqq a$, $0 \leqq y \leqq b)$, possessing continuous partial derivatives of all orders, mapping the closed region, S, on the closed rectangle, R, in a one-to-one manner and continuously, and such that $m = \phi(X, Y)$ and $n = \psi(X, Y)$. Let R' $(a' \leqq x \leqq a''$, $b' \leqq y \leqq b'')$ be any rectangle interior‡ to R and containing (X, Y) within its boundary Γ', and let L' be the curve in S corresponding to Γ'. Then the integral of $f(M + Ni)\cdot d(M + Ni)$ taken around L in the positive sense is equal to the integral of $f(M + Ni)\cdot d(M + Ni)$ taken in the positive sense around L',*

$$\int_L f(M + Ni)d(M + Ni) = \int_{L'} f(M + Ni)d(M + Ni).$$

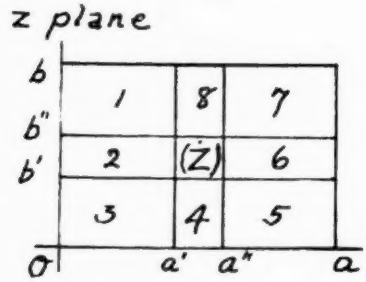


FIG. 6.

* O.C., p. 388 ff.

† O., p. 331 ff.; P., p. 206 ff.; G., p. 114 ff.

‡ I.e., $0 < a' < a'' < a$, $0 < b' < b'' < b$.

The proof follows immediately from the fundamental theorem I by application successively to the rectangles marked 1 ··· 8 in Fig. 6 and addition of the resulting equations. The equivalent of this equation in Cauchy's paper gives the first statement of his discovery on Residues.*

We proceed to derive the formulæ necessary to make the first application by evaluating $\int_{L'} f(M + Ni)d(M + Ni)$† explicitly for the case $M = x$, $N = y$, or $z = M + Ni$.

Let‡ $A + Bi = \int_{L'} f(z)dz$, and suppose:

Case 1.

$$f(z) = \frac{C}{z - Z}, \qquad \text{where} \qquad Z = X + Yi.$$

Then

$$A + Bi = \int_{L'} \frac{Cdz}{z - Z}.$$

Let

$$z - Z = re^{\phi i}, \qquad dz = ire^{\phi i}d\phi + e^{\phi i}dr.$$

Then

$$A + Bi = \int_0^{2\pi} Cid\phi + \int_{L'} \frac{dr}{r}$$

or

$$A + Bi = 2\pi iC,\text{§}$$

since the initial and final values of r are equal.

Case 2.

$$f(z) = \phi(z) + \frac{C}{z - Z},$$

where $\phi(z)$ is analytic in R and continuous in R and on the boundary, L, and where Z is within R,

$$A + Bi = \int_{L'} \phi(z)dz + \int_{L'} \frac{C}{z - Z}dz$$

or

$$A + Bi = 2\pi iC,$$

since $\int_{L'} \phi(z)dz = 0$ by the fundamental theorem I.

Case 3. If $f(z)$ has a pole on the boundary, L, of the rectangle, R, but

* O.C., p. 381, equation (4).

† L' is identical with Γ' in this case.

‡ The notation has here been changed from $A' + A''i$ to $A + Bi$.

§ Cf. O. C., footnote, p. 412, equation (C).

not at one of the vertices, we construct R'' as in figure 7 and denote by $\bar{L}$ and L'' respectively the boundaries of the rectangles R and R'', omitting in each case the segment AB. We apply now the fundamental theorem I to the rectangles 1, 2, 3 and sum the results. In this way we find

$$\int_{\bar{L}} f(z)dz + \int_{L''} f(z)dz = 0.$$

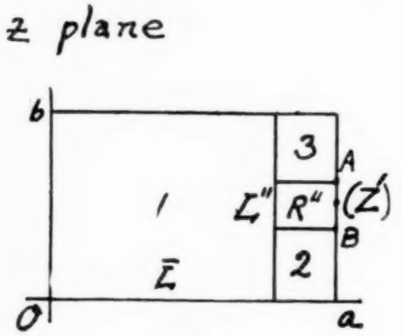


FIG. 7.

If now, in particular, we suppose $f(z) = C/(z - Z')$, we write $A + Bi = \int_{\bar{L}} f(z)dz = -\int_{L''} C/(z - Z')dz$, and $z - Z' = re^{\phi i}$; then $dz = ire^{\phi i}d\phi + e^{\phi i}dr$. Hence

$$A + Bi = \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} Cid\phi + \int_{L''} C\cdot\frac{dr}{r} = \pi iC,$$

since L'' may be chosen in such a manner that the initial and final values of r are equal.

Case 4. Let $f(z) = C/(z - Z') + \phi(z)$ where $\phi(z)$ is analytic throughout R, and Z' is a point of L, not at a vertex of R. Now $\int_L \phi(z)dz = 0$ by the fundamental theorem I. Hence here also we obtain

$$A + Bi = \pi iC.$$

Case 5. In general let us consider a rational function

$$f(z) = \frac{G(z)}{F(z)} = \sum\frac{C_k}{z - Z_k} + \sum\frac{C_k'}{z - Z_k'} + \phi(z),$$

where $F(z)$ has only simple roots in R, and on the boundary, L, *but not at a vertex*. Then $A + Bi$ must be computed for each pole and the results summed. Hence

(11) $$A + Bi = 2\pi i \sum C_k + \pi i \sum C_k'^{*}$$

where C_k is the coefficient of $1/(z - Z_k)$, Z_k an interior point of R, and where C_k' is the coefficient of $1/(z - Z_k')$, Z_k' a point on the boundary not at a vertex.

Let $C_k = \lambda_k - i\mu_k$ and $C_k' = \lambda_k' - i\mu_k'$. Then

(12) $$A = 2\pi \sum \mu_k + \pi \sum \mu_k'$$

and

(13) $$B = 2\pi \sum \lambda_k + \pi \sum \lambda_k'.$$

On the basis of the fundamental theorem II and equations (12) and (13) we may work out one of the examples given by Cauchy in Part II. Cauchy applies these formulæ to the rectangle bounded by $y = 0$, $y = b > 0$, $x = -a$, $x = a$, and then allows a and b to increase without limit (Fig. 8).

$$f(M + Ni) = f(z) = f(x + yi) = P + Qi,$$

where $Q(x, 0) \equiv 0$.

$$\begin{aligned} A + Bi &= \int_L f(z)dz \\ &= \int_{-a}^{a} P(x, 0)dx + \int_0^b [P(a, y) + iQ(a, y)]idy \\ &\quad - \int_{-a}^{a} [P(x, b) + iQ(x, b)]dx - \int_0^b [P(0, y) + iQ(0, y)]idy. \end{aligned}$$

Separating this equation into real and imaginary parts, we have

(14) $$A = \int_{-a}^{a} P(x, 0)dx - \int_0^b Q(a, y)dy - \int_{-a}^{a} P(x, b)dx + \int_0^b Q(0, y)dy,$$

(15) $$B = \int_0^b P(a, y)dy - \int_{-a}^{a} Q(x, b)dx - \int_0^b P(0, y)dy.$$

To be able to eliminate from the formulæ all integrals except those along the axis of reals, Cauchy thinks it sufficient to take $f(z)$ to be a function such that P and Q vanish when $x = \pm\infty$, $y = \infty$. This is not at all sufficient, however, for stronger conditions are called for to insure the vanishing of the integrals in question. It is sufficient, however, if a and b increase indefinitely in a prescribed manner such as e.g.

$$\lim_{a \to \infty} \frac{b}{a} = k \neq 0$$

and that

$$\lim_{a \to \infty} \sqrt{a^2 + b^2} \max |f(x + yi)| = 0,$$

* Cf. O.C., p. 422, footnote, equation (G).

the maximum being taken for all points $\pm a + yi$ in the interval $0 \leqq y \leqq b$ and all points $x + bi$ in the interval $-a \leqq x \leqq a$, i.e., for all points on three sides of the rectangle determined by $(-a, 0)$, $(-a, b)$, (a, b), and $(a, 0)$. That is, we shall assume that, given a positive number, ϵ, arbitrarily small, we can find a positive number, X, such that

$$\sqrt{a^2 + b^2} \max |f(x + yi)| < \epsilon,$$

when $a > X$ for all points $\pm a + yi$ in the interval $0 \leqq y \leqq b$ and all points $x + bi$ in the interval $-a \leqq x \leqq a$. Then

$$\left|\int_0^b Q(\pm a, y)dy\right| \leqq \int_0^b |f(\pm a + yi)|\, dy$$
$$\leqq \int_0^b \frac{\epsilon}{\sqrt{a^2 + b^2}} dy < \frac{\epsilon b}{\sqrt{a^2 + b^2}} < \epsilon,$$

when $a > X$.

Hence, as a and b increase indefinitely in this prescribed manner,

$$\lim_{a,\, b \to \infty} \int_0^b Q(\pm a, y)dy = 0.$$

Similarly, it may be proved that

$$\lim_{a,\, b \to \infty} \int_{-a}^{a} P(x, b)dx = 0, \quad \lim_{a,\, b \to \infty} \int_{-a}^{a} Q(x, b)dx = 0, \quad \lim_{b \to \infty} \int_0^b P(0, y)dy = 0,$$

$$\lim_{b \to \infty} \int_0^b Q(0, y)dy = 0, \qquad \lim_{a,\, b \to \infty} \int_0^b P(a, y)dy = 0.$$

Moreover, if in addition $\lim_{a \to \infty} af(a) = 0$, then $\int_{-a}^{a} P(x, 0)dx$ converges as a increases indefinitely, for if ϵ is positive and arbitrarily small, there exists a positive number X such that $|f(x)| < \epsilon/a$ for $a > x > X$, and

$$\left|\int_X^a P(x, 0)dx\right| \leqq \int_X^a |f(x)|\, dx$$
$$\leqq \int_X^a \frac{\epsilon}{a} dx$$
$$\leqq \frac{\epsilon}{a}(a - X) < \epsilon$$

for $a > X$.

Formula (15) tells us that, for a function fulfilling the above conditions, $B = 0$, and the formula (14) reduces to

$$\int_{-\infty}^{\infty} p dx = A,$$

where A is taken for all the poles of $f(z)$ where $y \geqq 0$. If $f(z)$ is an even function, (16) becomes

$$A = 2\int_0^\infty p\,dx.$$

Let $f(z) = z^{2m}/(1 + z^{2n})$, where n is the greater of the two positive integers, m and n. Then $Q(x, 0) \equiv 0$ and $P(x, 0) \equiv x^{2m}/(1 + x^{2n})$, an even function. Hence

$$A = 2\int_0^\infty \frac{x^{2m}}{x^{2n} + 1}\,dx. \tag{17}$$

We must now specify a path for $a + bi$ such that $\lim\limits_{a\to\infty} \dfrac{b}{a} = k \neq 0$ and $\lim\limits_{a\to\infty} \sqrt{a^2 + b^2}\max|f(x + yi)| = 0$ for all points $\pm a + yi$ in the interval $0 \leqq y \leqq b$ and all points $x + bi$ in the interval $-a \leqq x \leqq a$.

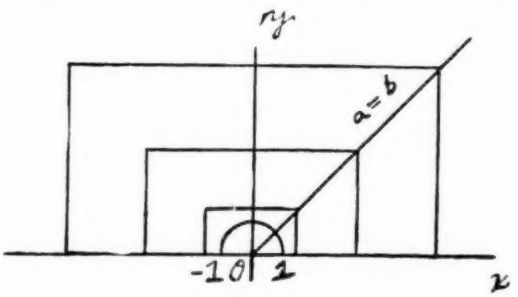


FIG. 8.

We will take $a = b$ (see Fig. 8). Then

$$\sqrt{a^2 + b^2}\max|f(\pm a + yi)| \leqq \sqrt{2}a\left|\frac{(\pm a + ai)^{2m}}{1 + (\pm a)^{2n}}\right|$$
$$\leqq \frac{\sqrt{2}\,|(\pm 1 + i)^{2m}|}{a^{2(n-m)-1}(a^{-2n} + 1)}$$
$$\leqq \frac{2^{2m+\frac{1}{2}}}{a^{-2n} + 1}\,\frac{1}{a^{2(n-m)-1}}.$$

The last expression approaches zero as a increases indefinitely.

Similarly

$$\sqrt{a^2 + b^2}\max|f(x + bi)| \leqq \sqrt{2}a\left|\frac{(a + ai)^{2m}}{1 + (ai)^{2n}}\right|$$
$$\leqq \sqrt{2}\left|\frac{(1 + i)^{2m}}{a^{-2n} + i^{2n}}\right|\frac{1}{a^{2(n-m)-1}}.$$

The last expression obviously approaches zero as a increases indefinitely. Also

$$\lim_{a\to\infty} af(a) = \lim_{a\to\infty}\frac{a^{2m+1}}{1 + a^{2n}} = 0.$$

The conditions sufficient to justify (17) are therefore satisfied.

The poles of $f(z)$ are to be found where $z^{2n} + 1 = 0$, or

$$Z_k = e^{\pi\left(\frac{2k+1}{2n}\right)i}, \qquad k = 0, 1 \cdots 2n-1.$$

We observe that the poles are on the unit circle and therefore certainly inside the rectangle R as soon as $a > 1$.

$$\frac{z^{2m}}{1+z^{2n}} = \frac{C_0}{z - e^{\frac{\pi}{2n}i}} + \frac{C_1}{z - e^{\frac{3\pi}{2n}i}} + \cdots + \frac{C_{2n-1}}{z - e^{\frac{4n-1}{2n}\pi i}}.$$

Now

$$\begin{aligned} C_k &= \lim_{z \to Z_k} \frac{z^{2m}(z - Z_k)}{1+z^{2n}} \\ &= \lim_{z \to Z_k} \frac{(2m+1)z^{2m} - 2mz^{2m-1}Z_k}{2nz^{2n-1}} \\ &= \frac{Z_k^{2m}}{2nZ_k^{2n-1}} = \frac{1}{2n} Z_k^{2(m-n)+1} \\ &= \frac{1}{2n} e^{\left[(2k+1)\frac{\pi i}{2n}\right](2m+1-2n)} \\ &= -\frac{1}{2n} e^{(2k+1)(2m+1)\frac{\pi i}{2n}}. \end{aligned}$$

Also

$$\begin{aligned} \sum_{k=0}^{2n-1} C_k &= -\frac{1}{2n} \sum_{k=0}^{2n-1} e^{(2k+1)\frac{2m+1}{2n}\pi i} \\ &= -\frac{1}{2n} \frac{1 - e^{(2m+1)\pi i}}{1 - e^{\frac{2m+1}{2n}\pi i}} e^{\frac{2m+1}{2n}\pi i} \\ &= -\frac{i}{2n} \frac{2i}{e^{\frac{2m+1}{n}\pi i} - e^{\frac{2m+1}{n}\pi i}} = -\frac{i}{2n\sin\frac{2m+1}{2n}\pi}. \end{aligned}$$

And

$$A = 2\pi \sum_{k=0}^{2n-1} \mu_k = \frac{2\pi}{2n\sin\frac{2m+1}{2n}\pi}.$$

Hence

$$\int_0^\infty \frac{x^{2m}}{1+x^{2n}}\,dx = \frac{\pi}{2n\sin\frac{2m+1}{2n}\pi}.$$

Let $2m + 1 = \alpha$ and $2n = \beta$. Then

$$\int_0^\infty \frac{x^{\alpha-1}}{1+x^\beta}\,dx = \frac{\pi}{\beta\sin\frac{\alpha}{\beta}\pi},$$

a formula which Euler had obtained.

In a similar manner other formulæ are obtained by taking other functions and other regions and integrating around the corresponding rectangle.*

We return to consider the method used by Cauchy in the second part to obtain the results of the fundamental theorem II and its immediate corollaries proved above. In the first place, Cauchy adopts the "principal value" definition to remove any difficulties regarding the evaluation of simple integrals due to singularities on the path of integration.†

Suppose we have $\phi'(x)$, the derivative of a real function $\phi(x)$ of a real variable. Consider

$$(19) \qquad \int_c^d \phi'(x)dx,$$

where $\phi'(x)$ has a finite or infinite discontinuity at (X) between c and d but is continuous in $c \leqq x < X$ and $X < x \leqq d$. By the "principal value" of (19) is meant

$$\begin{aligned}\lim_{h\to 0}\left[\int_c^{X-h} \phi'(x)dx + \int_{X+h}^d \phi'(x)dx\right] &= \lim_{h\to 0}\left[-\phi(c) + \phi(X-h) - \phi(X+h) + \phi(d)\right] \\ &= \phi(d) - \phi(c) - \Delta,\end{aligned}$$

where

$$\Delta = \lim_{h\to 0}\left[\phi(X+h) - \phi(X-h)\right].$$

According to the modern definition of convergence of an improper integral, the existence of this limit is a necessary but not a sufficient condition for the convergence of (19); Cauchy takes it as his working definition.‡ Secondly, to define the improper iterated integrals which occur, Cauchy proceeds as follows. Let $U(x, y)$ be a function which is continuous in x and y and possesses a continuous partial derivative with respect to x everywhere inside a rectangle, R $(0 \leqq x \leqq a, 0 \leqq y \leqq b)$, and on the boundary, L, except at the point $(0, 0)$ where it possesses a non-removable singularity but does not become infinite.§ Then‖

$$(20) \qquad \int_0^b dy \int_0^a \frac{\partial U}{\partial x}dx = \lim_{\xi\to 0}\int_0^b dy \int_\xi^a \frac{\partial U}{\partial x}dx$$

where $\xi > 0$.

This definition is totally inadequate, since the simple integral obtained after a first integration does not even come under the principal value definition and may even *diverge*.

* See O., pp. 289–295; G., pp. 118–122.

† O.C., p. 402.

‡ O.C. Cf. example on p. 404, $\int_{-2}^{4} dz/z$.

§ Cf. examples given by Cauchy, O.C., p. 394 and p. 397.

‖ O.C., p. 390.

It is, however, to be noticed that these insufficient definitions do not impair the value of Cauchy's results, nor do they substantially affect the method. If the discontinuity occurs at a corner of R, the method is not applicable in general, as Cauchy's formula itself shows.* For the definition of equation (20) above is an attempt to "cut-out" the singularity. But this does not yield a convergent result for this case. We cannot, therefore, put any real content into this particular result from our modern point of view and have, therefore, excluded it in our treatment. Cauchy, himself, makes no use of this equation in any of the numerous applications of the memoir.

When the point of discontinuity occurs inside of R at (X, Y), the rectangle is divided into four parts† by the lines $x = X$, $y = Y$, and the iterated integral is separated in a manner corresponding to the double integrals over each of the four rectangles. When these four integrals are added together, we have in rather obscure form what amounts to the method which we have set forth above. The singular point has been "cut-out" by a small rectangle, $x = X - \xi$, $x = X + \xi$, $y = Y - \eta$, $y = Y + \eta$, and a method of evaluating the line integral about the small rectangle is given in the form‡

$$(21) \qquad A = \lim_{\eta\to 0}\lim_{\xi\to 0}\left[\int_y^{y+\eta} U(X+\xi, y)dy + \int_{y-\eta}^{y} U(X+\xi, y)dy - \int_y^{y+\eta} U(X-\xi, y)dy - \int_{y-\eta}^{y} U(X-\xi, \xi)dy\right].$$

The bracket is substantially the real part of our equation (11). The difference between our method of evaluation of (11) and Cauchy's method of evaluating§ (21) is a striking example of the economy of the complex variable formulation.

When the point of discontinuity is on the boundary of R, the value of A is given by two terms‖ of (21). In both cases the results after integration are identical with those of (12) and (13).

In the historical review of the theory of functions by Brill and Noether¶ a brief treatment of the historical importance of the memoir is given but not from a critical point of view as is here done.**

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS.

* O.C. Cf. third equation under (20), p. 412.

† O.C., p. 396.

‡ O.C. Cf. p. 397, equation (13).

§ O.C., pp. 406–412.

‖ O.C., p. 400.

¶ Jahresbericht der deutschen Mathemataken-vereinigung, vol. 3 (1894), p. 165 ff.

** The above paper has grown out of an investigation of Cauchy's work on definite integrals and residues in a Seminar course at Harvard University. Some of the early work was done with the collaboration of Dr. E. S. Allen.

ARITHMETICAL DEDUCTION OF KRONECKER'S CLASS-NUMBER RELATIONS.

By G. H. Cresse.

The class-number relations which appear near the end of this paper are three of the eight celebrated class-number recursion formulas which L. Kronecker published* in 1860. In a preliminary announcement,† he had said, "If n denote an odd number > 3 and κ denote the modulus of an elliptic function, then the number of different values of κ^2 for which multiplication of the elliptic function by $\sqrt{-n}$ is possible is six times the number of classes of quadratic forms belonging to the determinant $-n$. Each value of κ^2 is the root of an integral equation whose degree is the number of such values of κ^2." Later‡ he intimated clearly that his only method of obtaining class-number relations from the theory of singular moduli was by setting two moduli equal to each other in the modular equation. By this method, H. J. Stephen Smith deduced§ in detail the eight formulas in a report which Kronecker has commended‖ for insight and mastery of principles.

C. Hermite¶ showed how a class-number relation can be obtained by equating the coefficients of like powers of $e^{\pi i\omega}$ in two expansions of a "doubly periodic function of the third kind." K. Petr** by Hermite's method deduced all eight of Kronecker's relations. In parallel researches, G. Humbert†† and L. J. Mordell‡‡ have reproduced independently many of Petr's intermediate results.

Kronecker§§ set up a one-to-one correspondence between certain quadratic forms and bilinear forms in four variables and then developed a theory of bilinear forms which established arithmetically the first six of his eight formulas. More interest however has been taken in the method of arithmetical deduction which was first illustrated by J. Liou-

* Jour. für Math., 57, 1860, 248–255; Jour. de math. (2), 5, 1860, 289–299.

† Monatsberichte Akad., Berlin, Oct., 1857, 456.

‡ Ibid., 1875, 235.

§ Report of the British Association, 35, 1865, 349–359.

‖ Monatsberichte Akad., Berlin, 1875, 234.

¶ Comptes Rendus, Paris, 53, 1861, 214–228; Jour. de math. (2), 7, 1862, 25–44; Œuvres, Paris, 1908, II, 109–124.

** Rozpravy ceske Akademia, Prague, 9, 1900, No. 38 (Bohemian language).

†† Jour. de math. (6), 3, 1907, 337–449.

‡‡ Messenger of Math., 45, 1916, 76–80.

§§ Monatsberichte Akad., Berlin, 1866, 873; Abhandlungen Königl. Preuss. Akad. Wiss., Berlin, 1883, II, 2d Abhand., pp. 60. Werke, Leipzig, 1897, II, 425–490.

ville* and which has been applied to a certain class-number relation by L. J. Mordell.† The method is the result of translating one of Hermite's analytic proofs into an arithmetical one. J. V. Uspensky‡ has accomplished the deduction of the eight formulas by this method. In the remainder of this paper, I shall reproduce the substance of his proof of formulas I, II, V and supply myself many desirable details including the proofs of his lemmas.

LEMMA 1. Let $F(x)$ be an uneven arithmetical function of the integer x; i.e., $F(-x) = -F(x)$, $F(0) = 0$. Let m be an uneven positive number and let

$$\sum F\left(\frac{d+\delta}{2}\right) = S, \tag{1}$$

where the summation extends to all integer solutions of the equation

$$m = 4h^2 + d\delta, \tag{2}$$

in which d and δ are positive but $h \gtreqless 0$. Let

$$\sum F(i + d') = S', \tag{3}$$

where the summation extends to all integer solutions of the equation

$$m = i^2 + 2d'\delta', \tag{4}$$

in which d' and δ' are positive and δ' uneven, but i is $\gtrless 0$. Then $S = 2S'$ if m is not a square, and $S = 2S' + \sqrt{m}\,F(\sqrt{m})$ if m is a square. That is,

$$\sum_{m=4h^2+d\delta} F\left(\frac{d+\delta}{2}\right) = 2\sum_{\substack{m=i^2+2\delta'd' \\ \delta'\equiv 1 \pmod 2}} F(i+d') + \begin{cases} 0, \text{ if } m \text{ is not a square,} \\ \sqrt{m}F(\sqrt{m}), \text{ if } m \text{ is a square.} \end{cases} \tag{A}$$

Proof§ of Lemma 1. Referring to (3) and (4), let

$$x = i + d', \qquad y = \delta' - i, \qquad z = i + d' - \delta';$$

that is

$$i = x - y - z, \qquad d' = y + z, \qquad \delta' = x - z.$$

Then (4) becomes all representations

$$m = x^2 + y^2 - z^2, \tag{5}$$

in which x and z are each even or uneven $\gtreqless 0$; y even $\gtreqless 0$; $y + z > 0$; $x > z$. But the sum (3) is not affected if we add the condition $x + z > 0$; for, corresponding to every solution (x, y, z) of (5) in which $x + z$ is < 0,

* Jour. de math. (2), 7, 1862, 44–48. Details of proof have been furnished by H. J. S. Smith, Report British Assoc., 35, 1865, 366–369; and by P. Bachmann, Niedere Zahlentheorie, Leipzig, 1910, II, 423–433.

† Messenger of Mathematics, 45, 1916, 177–180.

‡ Math. Sbornik, Moscow, 29, 1913, 26–52 (Russian language).

§ Cf. H. J. Stephen Smith, Rep. Brit. Assoc., 35, 1865, 368.

there is a solution $(-x, y, z)$ in which $x + z$ is < 0. Hence we consider in (5) all and only the representations in which

$$(6) \qquad y + z > 0, \qquad x > |z|.$$

That is to say, in (3), S' is equal to $\sum F(x)$ in which the summation extends to all solutions (x, y, z) of (5), (6).

Referring to (1) and (2), let

$$x = \frac{1}{2}(d + \delta), \qquad z = \frac{1}{2}(d - \delta), \qquad y = 2h.$$

Then (2) becomes all the representations

$$(7) \qquad m = x^2 + y^2 - z^2,$$

in which y is even and $\gtreqless 0$ and

$$(8) \qquad x > |z|.$$

That is to say, in (1), S is equal to $\sum F(x)$ in which the summation extends to all solutions (x, y, z) of (7), (8).

(*a*) *Case of m a non-square*, i.e., $|y| \neq |z|$. If $(x > 0, y > 0, z < y)$ is a solution of (5), (6), then $(x, y, -z)$ but neither $(x, -y, z)$ nor $(x, -y, -z)$ is a solution of (5), (6); while all four sets are solutions of (5), (6), while all four sets are solutions of (7), (8).

If $(x > 0, y > 0, z > y)$ is a solution of (5), (6), then $(x, -y, z)$ but neither $(x, y, -z)$ nor $(x, -y, -z)$ is a solution of (5), (6), while all four sets are solutions of (7), (8).

If $(x > 0, y < 0, z > y)$ is a solution of (5), (6), then $(x, -y, z)$ but neither $(x, y, -z)$ nor $(x, -y, -z)$ is a solution of (7), (8).

The categories given in the last three paragraphs of solutions of (7), (8) are exhaustive. Hence the lemma is proved for m a non-square.

(*b*) *Case of m a square*. To the solutions of (5), (6) for case (*a*), there will now be added only those solutions (x, y, z) in which $y = z > 0$. But corresponding to each of these new solutions of (5), (6) there are the new solutions (x, y, z), $(x, -y, z)$, $(x, y, -z)$, $(x, -y, -z)$ of (7), (8). The number of such solutions $(x, y, -z)$ and $(x, -y, z)$ combined with the solution $(x, 0, 0)$ of (7), (8) is $\sqrt{m}$ and the sum of the corresponding terms of (1) is $\sqrt{m}\, F(\sqrt{m})$. This completes the proof of *Lemma* 1.

Let $f(x)$ be an even function of the integer x. Let σ be an arbitrary real number. Then the function

$$F(x) = f(x - \sigma) - f(x + \sigma)$$

is an uneven function of (x). In (A), we replace m by $m - 2\rho\sigma$, where ρ and σ are given uneven positive numbers; and we take F as just defined. Then (A) becomes

$$\sum_{\substack{d,\,\delta>0\\ m=2\rho\sigma+4h^2+d\delta\\ h\gtreqless 0}} \left[f\left(\frac{d+\delta}{2}-\sigma\right)-f\left(\frac{d+\delta}{2}+\sigma\right)\right]$$

$$= 2 \sum_{\substack{i,\,d\\ i\gtreqless 0,\,d>0,\\ m=2\rho\sigma+i^2+2d\delta\\ \delta\equiv 1\,(\mathrm{mod}\ 2),\,>0}} [f(i+d-\sigma)-f(i+d+\sigma)] \tag{B}$$

$$+\left\{\begin{matrix} 0 \text{ if } m-2\rho\sigma \text{ is not a square}\\ s[f(s-\sigma)-f(s+\sigma)] \text{ if } m-2\rho\sigma=s^2>0\end{matrix}\right\}.$$

We take hereafter in this paper $m = 4n + 1$ and for this case the brace in the last equation is equal to zero, since $m - 2\rho\sigma = s^2$ has no solution for odd ρ and σ. We may now extend the summation in (B) to all possible ρ and σ, and have:

$$\sum_{\substack{d,\,\delta,\,\sigma>0\\ m=4h^2+2\rho\sigma+d\delta\\ h\gtreqless 0,\,\rho>0\\ \rho\equiv\sigma\equiv 1\,(\mathrm{mod}\ 2)}} \left[f\left(\frac{d+\delta}{2}-\sigma\right)+f\left(\frac{d+\delta}{2}+\sigma\right)\right]$$

$$= 2\cdot \sum_{\substack{i,\,d,\,\sigma>0\\ m=2\rho\sigma+i^2+2d\delta\\ \delta,\,\rho>0,\,\delta\equiv\rho\equiv\sigma\equiv 1\,(\mathrm{mod}\ 2)}} [f(d-\sigma+i) \tag{C}$$

$$+ f(d-\sigma-i) - f(d+\sigma+i) - f(d+\sigma-i)].$$

The right member of (C) is transformed by means* of

Lemma 2. Let n be an even number and consider all the representations

$$n = d\delta + \rho\sigma,$$

in which d, δ, ρ, σ are positive uneven numbers; also consider all the representations

$$n = d\delta,$$

in which d, δ are positive integers, δ uneven. If $\psi(x)$ is an even function, then

$$2\sum_{\substack{\rho,\,\sigma,\,d,\,\delta>0,\\ \equiv 1\,(\mathrm{mod}\ 2)\\ n=\rho\sigma+d\delta}} [\psi(d-\sigma)-\psi(d+\sigma)] = \sum_{\substack{d>0\\ n=d\delta\\ \delta>0,\,\equiv 1\,(\mathrm{mod}\ 2)}} d[\psi(0)-\psi(d)].$$

Proof† *of Lemma* 2. Consider the system

$$d\delta' + \rho'\sigma = n, \qquad d + \sigma = 2\mu, \qquad \delta' - \rho' = 2\nu; \tag{a}$$

in which d, δ' ρ', σ are positive and uneven, and μ, ν are given positive integers. The solutions of this system are equal in number to the solutions of the system

$$d\delta' + \rho'\sigma = n, \qquad d - \sigma = -2\mu, \qquad \delta' + \rho' = 2\nu. \tag{a'}$$

* Cf. J. Liouville, Jour. de math. (2), 3, 1858, 194.

† Cf. H. J. S. Smith, Rep. Brit. Assoc., 35, 1865, 366–367.

For, eliminating δ' and σ, we find that (a) has as many solutions as $n/2 = \nu d + \mu\rho'$ has solutions in which $d < 2\mu$ and (a') has as many as the same equation has solutions in which $\rho' < 2\nu$. These numbers of solutions are the same. For, corresponding to each solution of $n/2 = \nu d + \mu\rho'$ in which $d < 2\mu$ and $\rho' > 2\nu$, there is a solution of $n/2 = \nu(d + 2k\mu) + \mu(\rho' - 2k\nu)$ in which $\rho' - 2k\nu < 2\nu$; k being so chosen that $2k\nu$ is the largest multiple of 2ν that is $< \rho'$.

Similarly, the solutions of the two systems

$$d\delta' + \rho'\sigma = n, \qquad d + \sigma = 2\mu, \qquad \delta' - \rho' = -2\nu; \tag{b}$$

and

$$d\delta' + \sigma\rho' = n, \qquad d - \sigma = -2\mu, \qquad \delta' + \rho' = 2\nu, \tag{b'}$$

are equal in number. Also the number of solutions of each of the two systems

$$d\delta' + \rho'\sigma = n, \qquad d + \sigma = d'', \qquad \delta' = \rho' = \delta''; \tag{c}$$

and

$$d\delta' + \rho'\sigma = n, \qquad d = \sigma = \delta'', \qquad \delta' + \rho' = d'', \tag{c'}$$

for each pair of conjugate divisors d'', δ'' of n, is $\dfrac{d''}{2}$.

Hence if $\psi(x)$ is an even function, the enumeration of the solutions of (a), (a'), (b), (b'), (c), (c') gives the *Lemma* 2.

We specialize the even function $\psi(x)$ as

$$\psi(x) = f(x - i) + f(x + i)$$

in which $f(x)$ and i have the same meanings respectively as above. So the *Lemma* 2 applied to the right member of (C) becomes

$$2 \sum_{\substack{\iota,\, d,\, \sigma > 0 \\ d \equiv \delta \equiv \rho \equiv \sigma \equiv 1 \pmod 2 \\ \frac{m - \iota^2}{2} = \rho\sigma + d\delta}} [f(d - \sigma + i) + f(d - \sigma - i)$$

$$- f(d + \sigma + i) - f(d + \sigma - i)]$$

$$= \sum_{\substack{d,\, \delta,\, \iota > 0 \\ \delta \equiv 1 \pmod 2 \\ \frac{m - \iota^2}{2} = d\delta}} d[2f(i) - f(d + i) - f(d - i)].$$

And hence (C) becomes

$$\sum_{\substack{d,\, \delta,\, \rho,\, \sigma > 0 \\ \rho \equiv \sigma \equiv 1 \pmod 2 \\ m = 4h^2 + 2\rho\sigma + d\delta \\ h \gtreqless 0}} \left[f\left(\frac{d + \delta}{2} - \sigma\right) - f\left(\frac{d + \delta}{2} + \sigma\right)\right]$$

$$= \sum_{\substack{\iota,\, d > 0 \\ m = \iota^2 + 2d\delta \\ \delta > 0 \\ \equiv 1 \pmod 2}} d[2f(i) - f(d + i) - f(d - i)]. \tag{C'}$$

From (C'), Kronecker's classical formulas I, II, V now follow by taking $f(\pm 1) = 1, f(x) = 0$ if x^2 is not $= 1$.

We evaluate the left member of (C'). Only the first f has significance and that only for the argument $\frac{d+\delta}{2} - \sigma = \pm 1$. We denote the uneven number $\frac{d-\delta}{2}$ by τ and take first $\frac{d+\delta}{2} - \sigma = +1$. It follows that $-\sigma \leqq \tau \leqq \sigma$. Consequently, if we set

$$2u = \sigma - \tau, \qquad 2v = \sigma + \tau,$$

the numbers u and v will be $\geqq 0$ and of different parity. Then since $m = 4n + 1$, it is easily found, when we set $\rho = 2k - 1$, that the above equation, $m = 4h^2 + 2\rho\sigma + d\delta$, is equivalent to either of the following three equations: *

$$\begin{aligned} n - h^2 &= (k+u)(k+v) - k^2, \\ n - h^2 &= (u+k)(u+v) - u^2, \\ n - h^2 &= (u+v)(u+k) - u^2, \end{aligned} \tag{D}$$

$h \gtreqless 0,\ k \geqq 1,\ u \geqq 0,\ v \geqq 0,\ u + v \equiv 1 \pmod 2$.

It is evident that the complete number of solutions of each equation in (D) is 2 times the number of those solutions in which u is $< v$. Hence we confine our study to these latter solutions of (D). To each solution of (D_1) in which $k \leqq u < v$, there corresponds a quadratic form (A, B, C) in which

$$A = u + k, \qquad B = k, \qquad C = v + k,$$

of determinant $h^2 - n < 0$, whose coefficients satisfy the condition

$$A < C, \qquad B > 0, \qquad 2B \leqq A, \qquad A + C \equiv 1 \pmod 2.$$

To each of the solutions of (D_2) in which $u < k \leqq v$, there corresponds a quadratic form (A, B, C) in which

$$A = u + k, \qquad B = u, \qquad C = u + v,$$

of determinant $h^2 - n < 0$, whose coefficients satisfy the condition

$$A \leqq C, \qquad B \geqq 0, \qquad 2B < A, \qquad C \equiv 1 \pmod 2.$$

To each of the solutions of (D_3) in which $u < v < k$, there corresponds a quadratic form (A, B, C) in which

$$A = u + v, \qquad B = u, \qquad C = u + k,$$

of determinant $h^2 - n < 0$, whose coefficients satisfy the condition

$$A < C, \qquad B \geqq 0, \qquad 2B < A, \qquad A \equiv 1 \pmod 2.$$

Conversely, to an arbitrary form (A, B, C) of any of the three types

* Cf. C. Hermite, Jour. de math. (2), 7, 1862, 32; Œuvres, Paris, 2, 1908, 116.

just considered, there corresponds uniquely a solution u, v, k, of (D). Hence the number of solutions of the above equation $m = 4h^2 + 2\rho\sigma + d\delta$ is

$$4P + 2Q + 2R + 2S,$$

in which P, Q, R, S denote numbers of forms (A, B, C) of determinant $h^2 - n < 0$:

P, the number of those forms in which $A < C$, $B > 0$, $2B < A$ and one of the numbers A and C is uneven;

Q, the number of those forms in which $A < C$, $2B = A$ and one of the numbers A and C is uneven;

R, the number of those forms in which $A < C$, $B = 0$ and one of the numbers A and C is uneven;

S, the number of those forms in which $A = C$, $B \geqq 0$, $2B < A$, $A \equiv 1 \pmod 2$.

Similarly, if we take $\dfrac{d+\delta}{2} - \sigma = -1$ in (C), it is found that the number of solutions of $m = 4h^2 + 2\rho\sigma + d\delta$ is

$$4P + 2Q + 2T + 2R;$$

in which P, Q, R have the same meaning as before, and T denotes the number of forms (A, B, C) of determinant $h^2 - n < 0$, satisfying the conditions

$$A = C, \qquad B > 0, \qquad 2B < A, \qquad A \equiv 1 \pmod 2.$$

But $8P + 4Q + 4R + 4S$ is the quadruple of the number of uneven classes of determinant $h^2 - n$. Denoting by $F(\Delta)$ the number of such classes of determinant $-\Delta$, we find then that the left member of (C') has the value

$$4 \sum_{\leqq \sqrt{n}} F(n - h^2) + 2T - 2S.$$

Now $2T - 2S = 0$, except when $n - h^2$ is the square of an uneven number and for such a value of $n - h^2$, $2T - 2S = -2$. Hence the left member of (C'), by our choice of the function $f(x)$, has the value

$$4\sum_h F(n - h^2) - 2\sigma(n),$$

in which the summation extends to all integral values of $h (\gtreqless 0)$ whose squares are $\leqq n$, and $\sigma(n)$ denotes the number of all representations of n in the form

$$n = s^2 + h^2,$$

where s is uneven and positive.

We evaluate the right member of (C'), namely:

$$2\sum df(i) - \sum df(d - i) - \sum df(d + i),$$

$$m = i^2 + 2d\delta, \qquad d, i > 0, \qquad \delta \equiv 1 \pmod 2, > 0.$$

In view of our choice of the function $f(x)$, the significant terms in the first sum are all and only those in which $i = 1$; those in the second sum have $i = d \pm 1$; the terms in the third sum are all zero.

The significant terms of the first sum correspond respectively to the solutions of

$$m = 4n + 1 = 1 + 2d\delta,$$

that is, to the solutions of

$$n = d'\delta, \qquad d = 2d',$$

and therefore the first sum is

$$2^{\alpha+2}X(n), \tag{E}$$

where $X(n)$ denotes the sum of the uneven divisors of n and 2^{α} is the highest power of 2 contained in n.

The terms of the second sum correspond respectively to solutions of

$$m = 4n + 1 = (d \pm 1)^2 + 2d\delta = d(d \pm 2 + 2\delta) + 1;$$

that is, to the solutions of

$$n = d'(d' \pm 1 + \delta), \qquad d = 2d'.$$

Hence the second sum will be represented by

$$2 \sum_{n=\Delta\Delta';\ \Delta' \leq \Delta} \Delta' + 2 \sum_{n=\Delta\Delta';\ \Delta' < \Delta} \Delta', \tag{F}$$

where the summations are extended to positive integers Δ and Δ' of the same parity.

(a) Suppose that n is of the form $4r$. Then the sum in (E) has the value

$$\begin{aligned} 2^3[X(r) + (2^{\alpha} - 1)X(r)] &= 2^3[X(r) + (1 + 2 + 2^2 + \cdots 2^{\alpha-2})X(r)] \\ &= 2^3[X(r) + \Phi(r)], \end{aligned}$$

where $\Phi(r)$ denotes the sum of the divisors of r.

The total sum expressed in (F) is now

$$8\Theta(r) + 4\epsilon\sqrt{r},$$

where $\Theta(r)$ denotes the sum of the divisors of r which are $< \sqrt{r}$ and $\epsilon = 1$ or 0 according as r is or not a square. We write

$$4\Theta(r) = 4Z(r) - 4\Psi(r),$$

where $Z(r)$ denotes the sum of the divisors of r which are $> \sqrt{r}$ and $\Psi(r)$ is defined by the identity. Moreover, by definition

$$4\Theta(r) + 4e\sqrt{r} = 4\Phi(r) - 4Z(r).$$

Combining the last two identities, we have for (F) the expression

$$4[\Phi(r) - \Psi(r)].$$

Since $\sigma(4r) = 0$, (C') now implies Kronecker's first class-number relation

$$F(4r) + 2F(4r - 1^2) + 2F(4r - 2^2) + \cdots = 2X(r) + \Phi(r) + \Psi(r).$$

(*b*) Suppose that n is of the form $2s$, s uneven. The sum in (E) will be $8\Phi(s)$; the sum in (F) will be lacking. The arithmetical function $\sigma(2s)$ is double the excess* of the number of divisors of s which have the form $4k + 1$ over that of divisors which have the form $4k - 1$. When we denote this difference ($= \sum_{\delta / r}(-1)^{\frac{\delta-1}{2}}$) by $\varphi(s)$, (C') gives Kronecker's second class-number relation

$$F(2s) + 2F(2s - 1^2) + 2F(2s - 2^2) + \cdots = 2\Phi(s) + \varphi(s).$$

(*c*) Suppose that n is the uneven number s. The right member of (C') is now $2\Phi(s) + 2\Psi(s)$; and $\sigma(s) = \varphi(s)$. Hence (C') implies Kronecker's fifth class-number relation

$$F(s) + 2F(s - 1^2) + 2F(s - 2^2) + \cdots = \tfrac{1}{2}[\Phi(s) + \Psi(s) + \varphi(s)].$$

In a similar deduction of Kronecker's formulas III, IV and VI, the analog of the above *Lemma* 1 is the following for $m = 4n + 1$:

$$\sum_{m = 4h^2 + d\delta} (-1)^h F\left(\frac{d+\delta}{2}\right) = 2 \sum_{\substack{m = i^2 + 2d'\delta' \\ \delta' \equiv 1 \pmod 2}} (-1)^{\frac{i-1}{2} + \frac{\delta-1}{2}} F(i+d');$$

in which the denotations are the same as in *Lemma* 1. The analog of *Lemma* 2 is here

$$\sum_{\substack{\tau = \rho\sigma + d\delta \\ \rho,\, \sigma,\, d,\, \delta \equiv 1 \pmod 2}} (-1)^{\frac{\delta-1}{2}}[\Theta(d+\sigma) - \Theta(d-\sigma)] = \sum_{\substack{\tau = d\delta \\ \delta \equiv 1 \pmod 2}} (-1)^{\frac{\delta-1}{2}} d\Theta(2d),$$

in which the denotations are the same as in *Lemma* 2, except that $\Theta(x)$ is an arbitrary uneven function; $\Theta(0) = 0$; and τ is even.

By setting $\Theta(x) = f(x + i) - f(x - i)$, where $f(x)$ is an arbitrary even function of x, the following analog of (C') is obtained:

$$\sum_{\substack{\sigma,\, d,\, \delta > 0 \\ \equiv 1 \pmod 2 \\ m = 4h^2 + 2\rho\sigma + d\delta \\ h \gtreqless 0 \\ \rho \equiv 1 \pmod 2 \\ > 0}} (-1)^h \left[f\left(\frac{d+\delta}{2} - \sigma\right) - f\left(\frac{d+\delta}{2} + \sigma\right)\right]$$

$$= 2 \cdot \sum_{\substack{i,\, d,\, \delta > 0 \\ \delta \equiv 1 \pmod 2 \\ m = i^2 + 2d\delta}} (-1)^{\frac{i-1}{2} + \frac{\delta-1}{2}} d[f(2d - i) - f(2d + i)].$$

Two other similar pairs of lemmas lead respectively to Kronecker's formulas VII, VIII.

UNIVERSITY OF ARIZONA,
TUCSON, ARIZ.

*Cf. L. E. Dickson, History of the Theory of Numbers, vol. II, p. 235.

CYCLOTOMIC HEPTASECTION FOR THE PRIME 43.

By Pandit Oudh Upadhyaya.*

The problem of cyclotomic section has engaged the attention of many eminent mathematicians and solutions have been obtained by them for particular cases. The problem of the trisection and quartisection was completely solved by Cayley in a paper in which he also discussed the quinquisection but did not complete the solution. He once again took up the same problem in the proceedings of the London Mathematical Society in 1881 but was not able to complete the solution.

The problem of quinquisection was first solved by Rogers.† The same problem has very recently been solved by Burnside.‡ Towards the end of his paper he refers to the case of heptasection and says "I have carried the case $q = 7$ so far as to assure myself that it is not quite parallel with that of $q = 5$; a set of three simultaneous Diophantine equations occur, but they are not sufficient to ensure that the equations expressing the product of A's form a consistent multiplication table." In view of this statement it is believed that the problem of heptasection for the prime 43 has not been previously considered.

The object of this paper is to consider the problem of heptasection for the prime 43. All the details of calculation have been suppressed in order to save space, and only the final result is given.

Let a be an imaginary root of $x^{43} - 1 = 0$, and let us divide all the imaginary roots into 7 groups according to the following scheme:

$$
\begin{aligned}
A &= a + a^{42} + a^{37} + a^{6} + a^{36} + a^{7},\\
B &= a^{3} + a^{40} + a^{25} + a^{18} + a^{22} + a^{21},\\
C &= a^{9} + a^{34} + a^{32} + a^{11} + a^{23} + a^{20},\\
D &= a^{27} + a^{16} + a^{10} + a^{33} + a^{26} + a^{17},\\
E &= a^{38} + a^{5} + a^{30} + a^{13} + a^{35} + a^{8},\\
F &= a^{28} + a^{15} + a^{4} + a^{39} + a^{19} + a^{24},\\
G &= a^{41} + a^{2} + a^{12} + a^{31} + a^{14} + a^{29}.
\end{aligned}
$$

Calculating the elementary symmetric functions of these expressions we get:

$$\sum A = -1, \quad \sum AB = -18, \quad \sum ABC = 35, \quad \sum ABCD = 38,$$
$$\sum ABCDE = -104, \quad \sum ABCDEF = 7, \quad ABCDEFG = 49.$$

* Babu Shiva Prasad Gupta Research Scholar.

† Lond. Math. Soc. Proc., vol. 32 (1900–01), pp. 199–207.

‡ Lond. Math. Soc. Proc., vol. (2) 14, (1915), pp. 251–259.

The equation whose roots are A, B, C, D, E, F, G is therefore

$$\eta^7 + \eta^6 - 18\eta^5 - 35\eta^4 + 38\eta^3 + 104\eta^2 + 7\eta - 49 = 0.$$

Every root of this equation may be expressed as a rational integral function of any one assigned root; it is therefore an Abelian equation and can be solved by radicals.

I should like to mention that I have received a great amount of help in calculation from Pandit Shukdeo Chaube, Babu Brahmdeo Roy, Babu Raichand Bothera, and Sohan Lal Dugar.

SUMMATION OF A DOUBLE SERIES.*

By T. H. Gronwall.

It is the purpose of the present note to show that the series

$$(1) \qquad F(x, y) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{(m+n-2)!\,(m+n-1)!}{m!\,(m-1)!\,n!\,(n-1)!} x^{2m}y^{2n},$$

which occurs in a physical problem,† has its region of convergence defined by $|x| + |y| < 1$, and that its sum is

$$(2) \qquad F(x, y) = \frac{1}{2}\big[1 - x^2 - y^2 - \sqrt{(1 + x + y)(1 + x - y)(1 - x + y)(1 - x - y)}\big],$$

where that branch of the square root is to be taken which reduces to $+1$ at $x = y = 0$.

We begin by showing that our series converges absolutely and uniformly for $|x| + |y| \leqq 1 - \epsilon$, where ϵ is as small as we please. The binomial expansion of $(|x| + |y|)^{m+n}$ contains only non-negative terms, one of which is $\frac{(m+n)!}{m!\,n!}|x|^m|y|^n$; consequently

$$\frac{(m+n)!}{m!\,n!}|x|^m|y|^n \leqq (1-\epsilon)^{m+n},$$

and therefore also

$$\frac{(m+n-2)!}{(m-1)!\,(n-1)!}\cdot\frac{(m+n-1)!}{m!\,n!}|x|^{2m}|y|^{2n} \leqq \frac{1}{n}|x|\,|y|^2(1-\epsilon)^{2m+2n-3} \leqq (1-\epsilon)^{2m+2n}.$$

The series $\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(1-\epsilon)^{2m+2n}$ being obviously convergent, our statement is proved.

By a well-known theorem on power series, (1) may be differentiated term by term, and we obtain

$$(3) \qquad \frac{1}{2y}\frac{\partial F(x, y)}{\partial y} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{(m+n-2)!\,(m+n-1)!}{m!\,(m-1)!\,(n-1)!\,(n-1)!} x^{2m}y^{2n-2}$$

for $|x| + |y| < 1$.

* Read before the American Mathematical Society, Feb. 26, 1921.

† K. W. Lamson, "Reflection of radiation from an infinite series of equally spaced planes," Physical Review, ser. 2, vol. 17 (1921), pp. 624–625.

Now assume that

$$(4) \qquad |x| < \frac{1}{2\sqrt{2}}, \qquad |y| < \frac{1}{2\sqrt{2}};$$

then $|x| + |y| < 1$, so that (3) converges absolutely, and for every z on the circle $|z| = \frac{1}{2}$, we have

$$(5) \qquad \left|(x^2+z)\left(1+\frac{y^2}{z}\right)\right| \leqq (|x|^2+|z|)\left(1+\left|\frac{y^2}{z}\right|\right) < \left(\frac{1}{8}+\frac{1}{2}\right)\left(1+\frac{1}{4}\right) < 1,$$

$$(6) \qquad \left|(x^2+z)\frac{y^2}{z}\right| < \left(\frac{1}{8}+\frac{1}{2}\right)\cdot\frac{1}{4} < 1.$$

By (5), the equation

$$(7) \qquad z - (x^2+z)(y^2+z) = 0$$

has no root z on the circle $|z| = \frac{1}{2}$ when x and y satisfy (4). The roots of (7) being continuous functions of x and y, and reducing to 0 and 1 for $x = y = 0$, it follows that when (4) is satisfied, (7) has one root z_1 where $|z_1| < \frac{1}{2}$ and another z_2 where $|z_2| > \frac{1}{2}$. Solving (7), we find

$$(8) \qquad z_1 = \frac{1}{2}\left[1 - x^2 - y^2 - \sqrt{(1+x+y)(1+x-y)(1-x+y)(1-x-y)}\right],$$

that branch of the square root being taken which reduces to $+1$ at $x = y = 0$.

After these preliminaries, we use Cauchy's theorem on the binomial expansion of $(x^2+z)^{m+n-2}$ to obtain

$$\frac{(m+n-2)!}{(m-1)!\,(n-1)!}x^{2m-2} = \frac{1}{2\pi i}\int_{|z|=\frac{1}{2}}(x^2+z)^{m+n-2}\frac{dz}{z^n},$$

and consequently

$$\sum_{\substack{m+n=k\\ m\geq 1,\, n\geq 1}} \frac{(m+n-2)!\,(m+n-1)!}{m!\,(m-1)!\,(n-1)!\,(n-1)!}x^{2m}y^{2n-2}$$
$$= \frac{1}{2\pi i}\int_{|z|=\frac{1}{2}}\frac{x^2(x^2+z)^{k-2}}{z}\sum_{n=1}^{k-1}\frac{(k-1)!}{(k-n)!\,(n-1)!}\frac{y^{2n-2}}{z^{n-1}}dz$$
$$= \frac{1}{2\pi i}\int_{|z|=\frac{1}{2}}\frac{x^2(x^2+z)^{k-2}}{z}\left[\left(1+\frac{y^2}{z}\right)^{k-1} - \left(\frac{y^2}{z}\right)^{k-1}\right]dz.$$

Since (3) is absolutely convergent by (4), it follows that

$$\frac{1}{2y}\frac{\partial F(x, y)}{\partial y} = \sum_{k=2}^{\infty}\frac{1}{2\pi i}\int_{|z|=\frac{1}{2}} \frac{x^2(x^2+z)^{k-2}}{z}\left[\left(1+\frac{y^2}{z}\right)^{k-1} - \left(\frac{y^2}{z}\right)^{k-1}\right]dz,$$

and by (5) and (6), summation and integration may be interchanged; summing the two geometric series thus obtained under the integral sign, we find

$$\frac{1}{2y}\frac{\partial F(x, y)}{\partial y} = \frac{1}{2\pi i}\int_{|z|=\frac{1}{2}} \frac{x^2}{z}\left[\frac{y^2+z}{z-(x^2+z)(y^2+z)} - \frac{y^2}{z-y^2(x^2+z)}\right]dz.$$

Assuming for the moment that $x \neq 0$, $y \neq 0$, it is seen that the residues of the integrand are

$$\begin{array}{lcl} 0 & \text{at} & z = 0, \\ \dfrac{x^2(y^2+z_1)}{z_1(1-x^2-y^2-2z_1)} & \text{"} & z = z_1, \\ -1 & \text{"} & z = \dfrac{x^2y^2}{1-y^2}, \end{array}$$

these being the only poles inside the circle of integration. Hence, by Cauchy's theorem

$$\begin{aligned}\frac{1}{2y}\frac{\partial F(x, y)}{\partial y} &= \frac{x^2(y^2+z_1)}{z_1(1-x^2-y^2-2z_1)} - 1 \\ &= \frac{x^2z_1 - [z_1-(x^2+z_1)(y^2+z_1)] + z_1^2}{z_1(1-x^2-y^2-2z_1)}\end{aligned}$$

or, since the bracket vanishes by (7),

$$\frac{1}{2y}\frac{\partial F(x, y)}{\partial y} = \frac{x^2+z_1}{1-x^2-y^2-2z_1},$$

or finally, calculating $\partial z_1/\partial y$ from (7),

$$\frac{\partial F(x, y)}{\partial y} = \frac{\partial z_1}{\partial y}. \tag{9}$$

This, being established for $0 < |x| < \dfrac{1}{2\sqrt{2}}$, $0 < |y| < \dfrac{1}{2\sqrt{2}}$, also holds for $|x| + |y| < 1$, both members being holomorphic in the latter region. Integrating (9) in respect to y, and observing that both $F(x, y)$ and z_1 vanish for $y = 0$, it follows that $F(x, y) = z_1$, so that (2) is proved for $|x| + |y| < 1$. Now suppose that the series (1) converges for $x = x_0$, $y = y_0$, where $|x_0| > 0$, $|y_0| > 0$. Then, as is well known, the series converges uniformly for all x, y satisfying the inequalities $|x| \leqq \rho|x_0|$, $|y| \leqq \rho|y_0|$, where ρ is any constant less than unity; therefore $F(x, y)$ is

holomorphic for all such values, in particular for $x = \rho|x_0|$, $y = \rho|y_0|$. Assuming $|x_0| + |y_0| > 1$, we may take $\rho = \dfrac{1}{|x_0| + |y_0|}$ and $F(x, y)$ would be holomorphic at a point x, y where $x + y = 1$, which is impossible by (2). Hence our series diverges when $|x| + |y| > 1$, unless either $x = 0$ or $y = 0$, in which case every term vanishes and the series converges. Whether it converges or diverges for $|x| + |y| = 1$ is left undecided.

ON THE POSITIONS OF THE IMAGINARY POINTS OF INFLEXION AND CRITIC CENTERS OF A REAL CUBIC.

By B. M. Turner.

1. **Introduction.** In the extensive study of the configuration formed by the points of inflexion of a real cubic, it appears that no one has considered the possible positions of the six imaginary points of the group when the three real points are fixed. This is worthy of consideration for these two sets of points are so related that, while the three collinear real points of inflexion impose only five conditions and hence determine a fourfold infinite system of cubics in a plane, not one of the six points can be chosen arbitrarily. The following gives a construction for such a set of six points when the three real points are taken arbitrarily on a line; and by a generalization accounts for all such possible sets of six points.

The construction for the six imaginary points of inflexion also serves to show the positions of the twelve critic centers for the non-singular real cubic.

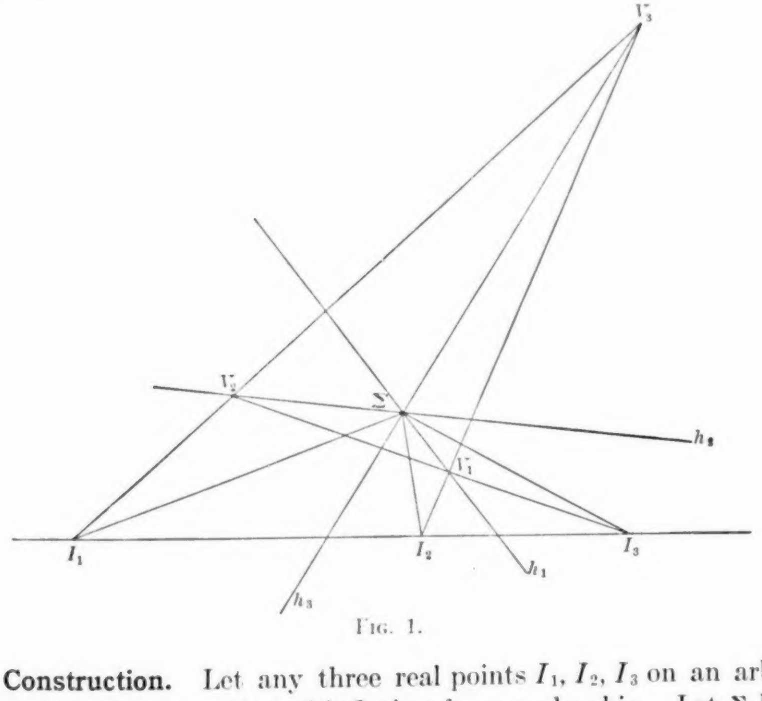


Fig. 1.

2. **Construction.** Let any three real points I_1, I_2, I_3 on an arbitrary real line be taken as points of inflexion for a real cubic. Let Σ be any

other real point. Join Σ to the points I_i and construct the fourth harmonic to each one of these lines with respect to the other two. Denote the fourth harmonic to the line through I_1 by h_1, and similarly for I_2 and I_3. Through any one of the three points, say I_1, draw an arbitrary real line intersecting h_2 and h_3 in V_2 and V_3. Draw the lines I_2V_3, I_3V_2 intersecting in V_1 on h_1. The projections upon the sides of the triangle V_1, V_2, V_3, through Σ, of the two points equianharmonic to the three points I, are six imaginary points which together with I_i form an inflexional group for a real cubic.

3. **Analytical proof of the construction.** Let $I_1(0,\ 1,\ -1)$, $I_2(-1,\ 0,\ 1)$, $I_3(1,\ -1,\ 0)$ be the three collinear points and $\Sigma(1,\ 1,\ 1)$ the arbitrary point of the plane. Then the lines joining Σ to I_i are

$$-2x + y + z = 0, \qquad x - 2y + z = 0, \qquad x + y - 2z = 0;$$

and the lines h_i are

$$h_1 : y - z = 0, \qquad h_2 : z - x = 0, \qquad h_3 : x - y = 0.$$

An arbitrary line through I_1 is $\alpha x + y + z = 0$, where α is an undetermined real number. This line cuts h_2 and h_3 in $V_2(-1,\ \alpha + 1,\ -1)$, $V_3(-1,\ -1,\ \alpha + 1)$, respectively. The lines I_2V_3, I_3V_2 have equations

$$x + \alpha y + z = 0, \qquad x + y + \alpha z = 0$$

and intersect on h_1 in $V_1(\alpha + 1,\ -1,\ -1)$.

The two points equianharmonic to I_1, I_2, I_3 are $(1,\ \omega,\ \omega^2)$, $(1,\ \omega^2,\ \omega)$, where $\omega^3 = 1$; and the lines joining these to Σ are

$$x + \omega y + \omega^2 z = 0, \qquad x + \omega^2 y + \omega z = 0.$$

These two lines intersect the sides of the triangle $V_1\ V_2\ V_3$ in

$$\begin{array}{ll}
(\omega^2 - \omega,\ 1 - \alpha\omega^2,\ \alpha\omega - 1), & (\omega - \omega^2,\ 1 - \alpha\omega,\ \alpha\omega^2 - 1); \\
(\alpha\omega - 1,\ \omega^2 - \omega,\ \alpha\omega^2 - 1), & (\alpha\omega^2 - 1,\ \omega - \omega^2,\ 1 - \alpha\omega); \\
(1 - \alpha\omega^2,\ \alpha\omega - 1,\ \omega^2 - \omega), & (1 - \alpha\omega,\ \alpha\omega^2 - 1,\ \omega - \omega^2).
\end{array}$$

The six points just determined together with I_i may be arranged in the scheme

$$\begin{array}{lll}
(0,\ 1, -1), & (\omega^2 - \omega,\ 1 - \alpha\omega^2,\ \alpha\omega - 1), & (\omega - \omega^2,\ 1 - \alpha\omega,\ \alpha\omega^2 - 1), \\
(\alpha\omega^2 - 1,\ \omega - \omega^2,\ \alpha\omega - 1), & (-1,\ 0,\ 1), & (\alpha\omega - 1,\ \omega^2 - \omega,\ 1 - \alpha\omega^2), \\
(1 - \alpha\omega^2,\ \alpha\omega - 1,\ \omega^2 - \omega), & (1 - \alpha\omega,\ \alpha\omega^2 - 1,\ \omega - \omega^2), & (1,\ -1,\ 0),
\end{array}$$

whose rows, columns, right and left hand diagonals satisfy the conditions of collinearity imposed on the nine points of inflexion of a cubic. Then from the scheme, for every value of α,

$$(x + y + z)(x + \omega y + \omega^2 z)(x + \omega^2 y + \omega z) \\ + \lambda(\alpha x + y + z)(x + \alpha y + z)(x + y + \alpha z) = 0$$

can be read off as the equation of the pencil of cubics with inflexions at the nine points.

For this pencil the lines

$$\alpha x + y + z = 0, \qquad x + \alpha y + z = 0, \qquad x + y + \alpha z = 0$$

are the sides of the real inflexional triangle; and Σ is the point common to the three real harmonic polars h_i. Hence the result may be stated in the theorem:

The six imaginary points of inflexion of a real cubic are the projections, through the point common to the three real harmonic polars, of the two points equianharmonic to the three real inflexions, upon the sides of the real inflexional triangle.

4. **Generalization.** The value of α depends upon the choice of the line through one of the points I, hence the equation

$$(x + y + z)(x + \omega y + \omega^2 z)(x + \omega^2 y + \omega z) \\ + \lambda(\alpha x + y + z)(x + \alpha y + z)(x + y + \alpha z) = 0,$$

depending upon two variable parameters, accounts for a two-fold infinite system of cubics—a syzygetic pencil for each of the single infinity of choices of the line with a given Σ. The double infinity of choices of Σ accounts for the fourfold infinite system of cubics in a plane with the same three real points of inflexion.

As Σ varies in position in the plane, the projections through it of the two equianharmonic points define the totality of imaginary points on the lines of the three real pencils

$$\alpha x + y + z = 0, \qquad x + \alpha y + z = 0, \qquad x + y + \alpha z = 0;$$

that is, every one of these points belongs to at least one inflexional group which includes the three given real points. On the other hand, an imaginary point on a real line not included in the three pencils cannot belong to such an inflexional group, and hence the impossibility of an arbitrary choice of an imaginary point of inflexion for a real cubic when the three real points of inflexion are fixed.

Thus is developed the following theorem:

The imaginary points of inflexion of the fourfold infinite system of real cubics in a plane, with three given real points of inflexion, form the totality of imaginary points on the three pencils of real lines through the three fixed inflexions; and group themselves into ∞^3 sets of six points, two points on one line of each pencil, such that each set of six together with the three fixed real points form an inflexional group.*

* The first half of this theorem was also proved in a former paper by the writer.

Two special cases arising when the arbitrarily chosen line is taken (1) through Σ, giving rational cubics with a conjugate point, and (2) coincident with the line through the three points I, giving degenerate cubics, have been considered in another connection in an earlier paper.

Since three collinear points do not determine a plane, it may further be noted that Σ may be taken as any real point in three-dimensional space, and the theorem extended accordingly.

5. **The critic centers.** It has been noted that the removal of the restriction that Σ be a fixed point gives the system of cubics in the plane two more degrees of freedom. A fixed Σ is a critic center (vertex of an inflexional triangle) for every cubic of the doubly infinite system. This suggests that, with no other restriction, four critic centers chosen arbitrarily in the plane may impose eight conditions and hence determine a singly infinite system of cubics.

Suppose the four points $(1, \pm 1, \pm 1)$ to be critic centers for a real cubic. Since the points are all real, three must be taken as vertices of the real inflexional triangle, say $(1, 1, -1)$, $(-1, 1, 1)$, $(1, -1, 1)$. The cubic consisting of the three sides of the triangle is

$$(y + z)(z + x)(x + y) = 0;$$

and the polar line of the fourth point $(1, 1, 1)$ with respect to this cubic is $x + y + z = 0$. On this line $(1, \omega, \omega^2)$, $(1, \omega^2, \omega)$ are the two, and the only two, points whose polar lines pass through $(1, 1, 1)$. Hence under the hypothesis $(1, 1, 1)$, $(1, \omega, \omega^2)$, $(1, \omega^2, \omega)$ are the vertices of a second syzygetic triangle which forms the cubic

$$(x + y + z)(x + \omega y + \omega^2 z)(x + \omega^2 y + \omega z) = 0,$$

and

$$(x + y + z)(x + \omega y + \omega^2 z)(x + \omega^2 y + \omega z) + \lambda(y + z)(z + x)(x + y) = 0$$

is a pencil of cubics with the four chosen points $(1, \pm 1, \pm 1)$ as critic centers. This is identical with the equation of the preceding section when α is equal to zero; hence the hypothesis holds, that is, four real points may be arbitrarily chosen in a plane as critic centers for a cubic. From among the four points there are four choices of three, and any such three may be taken as the vertices of the real inflexional triangle. This gives the theorem:

Four real points chosen arbitrarily in a plane as critic centers for a real cubic determine a syzygetic pencil of cubics as one of four.

The nine points of inflexion for the pencil

$$(x + y + z)(x + \omega y + \omega^2 z)(x + \omega^2 y + \omega z) + \lambda(y + z)(z + x)(x + y) = 0$$

are

$$(0,\ 1,\ -1),\qquad (\omega^2-\omega,\ 1,\ -1),\qquad (\omega-\omega^2,\ 1,\ -1),$$
$$(-1,\ \omega-\omega^2,\ 1),\qquad (-1,\ 0,\ 1),\qquad (-1,\ \omega^2-\omega,\ 1),$$
$$(1,\ -1,\ \omega^2-\omega),\qquad (1,\ -1,\ \omega-\omega^2),\qquad (1,\ -1,\ 0),$$

which define the sides of the two other inflexional triangles and consequently the remaining six critic centers as

$$(2\omega-1,\ 1,\ 1),\qquad (2\omega^2-1,\ 1,\ 1);$$
$$(1,\ 2\omega-1,\ 1),\qquad (1,\ 2\omega^2-1,\ 1);$$
$$(1,\ 1,\ 2\omega-1),\qquad (1,\ 1,\ 2\omega^2-1).$$

The six points just determined are the intersections of

$$y-z=0,\qquad z-x=0,\qquad x-y=0,$$

by the lines joining $(1, 1, -1)$, $(-1, 1, 1)$, $(1, -1, 1)$ to the two points $(1, \omega, \omega^2)$, $(1, \omega^2, \omega)$; that is, they are the projections on the three harmonic polars, through the vertices of the real inflexional triangle, of the two points equianharmonic to the three real inflexions.

Then, in the figure, the twelve critic centers are V_1; V_2; V_3; Σ; the two points equianharmonic to I_1, I_2, I_3; and the projections upon h_i, through V_i, of the two equianharmonic points.

University of Illinois,
1921.

FREQUENCY DISTRIBUTIONS OBTAINED BY CERTAIN TRANSFORMATIONS OF NORMALLY DISTRIBUTED VARIATES.*

BY H. L. RIETZ.

The problem considered in this paper was first suggested to the writer by experiments with actual frequency distributions of various measurements of objects which approximate roughly to a set of similar solids. To be concrete, we may think of the diameters, surfaces, and volumes of spheres that represent objects in nature, such as oranges on a tree or peas on a plant.

Suppose the distribution of diameters is a normal distribution given by

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\bar{x})^2}{2\sigma^2}}.$$

It seems natural to inquire into the nature of the distribution of the corresponding surfaces and volumes. Conversely, we should ask for a determination of the distribution of diameters if we knew surfaces or volumes were normally distributed. The same kind of problem† would arise if we knew that velocities, v, of molecules of a gas were normally distributed, and were required to investigate the distribution of energy $\frac{1}{2}mv^2$. These concrete illustrations are special cases of the transformation of variates of a normal distribution by replacing each variate, x, by an assigned function kx^n, where k is a positive constant and n is a positive integer or the reciprocal of a positive integer. Edgeworth‡ and Kapteyn§ have made use of transformations of the normal curve as a method of representing skew frequency distributions. Apart from the possible use for this purpose, the frequency curves arising from certain simple transformations of the variates of a normal distribution present points of special interest to which it seems that attention should be directed, particularly because of the striking differences in general appearance from normal curves—a fact that seems both interesting and important in forming a proper conception of the place of the normal curve in the representation of frequency distributions.

It is the main purpose of the present paper to exhibit certain properties

* Read before the American Mathematical Society at Lincoln, Nebraska, Nov. 27, 1920.

† Edgeworth, Proc. Fifth International Congress of Mathematicians, II, p. 427.

‡ Loc. cit. and a series of papers in the Journal of the Royal Statistical Society. See vol. 61, pp. 670–700.

§ Skew Frequency Curves in Biology and Statistics, 1903.

of the frequency curves that are obtained when the variates of a normal distribution are transformed by substituting for each variate, x, the function kx^n where k is a positive constant, and where suitable restrictions will be placed on n as we proceed. The case $n = 1$ is treated by Bruns* and the results are simple and well known. Edgeworth called attention to the general form of the frequency curve with which we are concerned for $n = 2$. Furthermore, when deviations of variates from their mean value are small compared to their mean value, it is well known that the distributions of squares and cubes of variates approach normal distributions sufficiently near for certain purposes. But in certain important statistical applications, the deviations of variates from their mean value cannot be reasonably regarded as small compared to the mean value. This latter class of distributions gives special importance to our problem.

When $k = 1$, the problem is that of exhibiting the properties of the frequency distribution of the nth powers of a set of normally distributed variates. This case seems to include essentially the same points of interest contained in the more general problem, since the transformation $x' = x^n$, followed by the linear transformation $x'' = kx'$, produces the same result as the transformation $x'' = kx^n$. Hence we shall in what follows deal with the transformation $x' = x^n$.

To determine the frequency function obtained by the transformation, let $x_1, x_2, \cdots, x_t$ be a system of variates expressed in a unit equal to the standard deviation σ, and belonging to the normal distribution

$$y = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-\bar{x})^2}{2}}, \qquad \bar{x} \gtreqless 0,$$

so that $P = \int_a^b y dx$ is the probability that a variate taken at random belongs to the interval a to b.

Let us replace each variate x_s by x_s', where $x_s' = x_s^n$. We then make a corresponding transformation of the integral $\int_a^b y dx$ by letting $x' = x^n$ $(n \neq 0)$. Then

$$dx' = nx^{n-1}dx, \text{ except at } x = 0 \text{ when } n < 1,$$

and

$$dx = \frac{dx'}{nx'^{\frac{n-1}{n}}}, \text{ except at } x' = 0 \text{ when } n > 1.$$

* Wahrscheinlichkeitsrechnung und Kollektivmasslehre, 1906, pp. 126–129.

We may therefore write

$$P = \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-(x-\bar{x})^2}{2}}\, dx = \frac{1}{n\sqrt{2\pi}} \int_{a^n}^{b^n} \frac{e^{-\frac{\left(x'^{\frac{1}{n}}-\bar{x}\right)^2}{2}}}{x'^{\frac{n-1}{n}}}\, dx', \tag{1}$$

where for the present we assume $a \geqq 0$, and $b \geqq 0$. As shown below, these limitations on a and b may be removed to some extent for certain values of n.

The frequency curve of x'-variates obtained from positive x-variates is then given by

$$y' = \frac{1}{n\sqrt{2\pi}\, x'^{\frac{n-1}{n}}} e^{-\frac{\left(x'^{\frac{1}{n}}-\bar{x}\right)^2}{2}}. \tag{2}$$

The function (2) does not represent a normal curve when $n \neq 1$.

In order to determine the general character of the frequency curve given by (2), we first examine the function for maxima and minima. For this purpose, we have

$$\frac{dy'}{dx'} = \frac{1}{n\sqrt{2\pi}} e^{-\frac{\left(x'^{\frac{1}{n}}-\bar{x}\right)^2}{2}} \left[-\frac{n-1}{n} x'^{-\frac{2n-1}{n}} - \frac{1}{n}(x'^{\frac{1}{n}} - \bar{x}) x'^{\frac{2-2n}{n}} \right].$$

The derivative changes signs at

$$x' = \frac{1}{2^n}(\bar{x} \pm \sqrt{\bar{x}^2 - 4(n-1)})^n, \tag{3}$$

when $\bar{x}^2 > 4(n-1)$, and at $x' = 0$ for certain values of n.

In equation (1) we restricted a and b to be zero or positive, but when n is an odd positive integer or the reciprocal of an odd positive integer, it follows at once that (2) gives the frequency curve corresponding to negative values of x' that arise from the transformation $x' = x^n$ when x is negative. By taking $\bar{x}$ sufficiently large, the function (2) may be made as nearly zero as we please for negative values of x' except at points near the discontinuity at $x' = 0$. This discontinuity exists when $n > 1$. When n is an odd number or the reciprocal of an odd number, the derivative dy'/dx' changes sign at $x' = 0$. When n is the reciprocal of an odd positive integer, there is a minimum at $x' = 0$, and the value of the function is zero at this minimum.

We shall find it convenient to consider the frequency curves given by (2) under three cases according as $n > 1$, $0 < n < 1$, or $n < 0$. We shall

limit our discussion to positive values of x and x' except when there is a specific statement extending the treatment to negative values.

Case I. $n > 1$.

The maximal frequency corresponds to the value of x' given by taking the positive sign before the radical in (3). In the language of statistics, the abscissa of this maximal frequency is called the modal value or the mode. We shall find it convenient to use these expressions later in this paper. The curve for $n = 3$, $\bar{x} = 4$ is shown in Fig. 1 for positive values of x'.

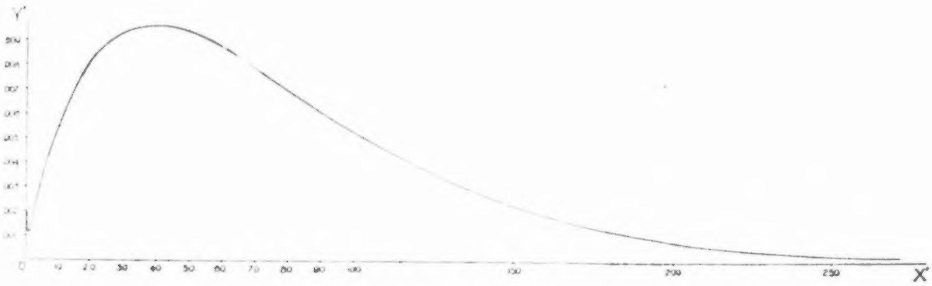


Fig. 1. $y' = \dfrac{x'^{-2/3}}{3\sqrt{2\pi}} e^{-\frac{(x'^{1/3}-4)^2}{2}}, \quad x' > 0.$

The skew appearance of the figure shows that the distribution is not even approximately a normal distribution. Since variates at the median of the original normal distribution must be transformed to the median of the new distribution, we may appropriately compare the value $\bar{x}^n$ of the median of the new distribution with the modal value given by (3). When $n > 1$, it follows from (3) that the mode of the new distribution is less than its median. The minimum that corresponds to the value of x' given by taking the negative sign before the radical in (3) is of special interest because the existence of this minimum would probably not be expected by ordinary intuition. Thus, when $\bar{x}^2 > 4(n - 1)$, we have a minimum between the origin and the maximum discussed above. This minimum is shown in Fig. 1 at a point near the origin for the case $n = 3$, $\bar{x} = 4$. The descent of the curve from infinity at $x' = 0$ to the minimum at $x' = (2 - \sqrt{2})^3$ is so rapid and the curve is so near the Y-axis that it cannot be shown well on the scale of Fig. 1. For this reason we show the curve in the neighborhood of this minimum in Fig. 2 on an enlarged scale. The function y' (n an odd number) has real positive values when x' is negative, but the values differ very little from zero, except near $x' = 0$, when $\bar{x} \geqq 4$. For the case $\bar{x} = 4$ shown in Fig. 1, the values of y' for

negative values of x' are so near zero, except at points near the discontinuity at $x' = 0$, that it is impractical to exhibit the curve on the scale of Fig. 1 for negative values of x'.

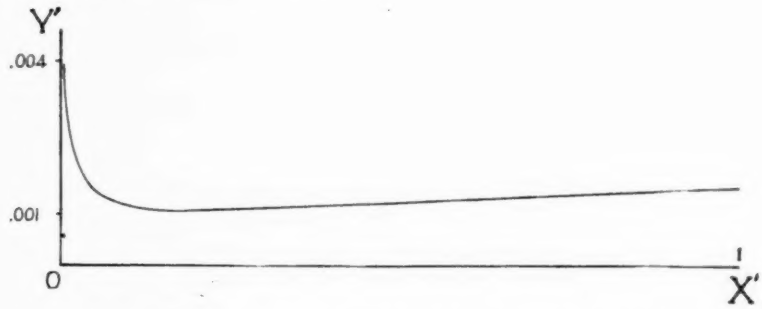


FIG. 2. Portion of curve of Fig. 1 from $x' = 0$ to $x' = 1$ with enlarged horizontal scale.

Next we find d^2y'/dx'^2. It turns out that the points of inflection are given by the solutions of the equation

$$x'^{\frac{1}{n}-3}[x'^{\frac{4}{n}} - 2\bar{x}x'^{\frac{3}{n}} + (3n - 4 + \bar{x}^2)x'^{\frac{2}{n}}$$
$$+ \bar{x}(3 - 3n)x'^{\frac{1}{n}} + (n - 1)(2n - 1)] = 0. \quad (4)$$

When $n = 3$, $\bar{x} = 4$, one point of inflection is at $x' = 1$ as shown in Fig. 1. Furthermore, it is easily verified when $\bar{x} = n + 1$, that (4) has a solution $x' = 1$, and that there is a point of inflection at $x' = 1$. There is also, in general, another point of inflection on the curve to the right of the maximum.

The general appearance of the frequency curve (2) depends much on the value of $\bar{x}$ compared to $4(n - 1)$. When $\bar{x}^2 \overline{\gtrless} 4(n - 1)$, the function (2) has no maximum nor minimum, but is a monotone decreasing function of x'. When $\bar{x}^2 = 4(n - 1)$, there is a point of inflection at $x' = \bar{x}^n/2^n$.

The problem when $n = 2$ and $\bar{x} > 2$ presents a point of special interest. Thus, if $\bar{x}$ were assigned larger and larger values, the x-coördinate of the minimum would approach zero and that of the maximum would approach the median $\bar{x}^2$. This is seen from the fact that

$$\tfrac{1}{4}(\bar{x} - \sqrt{\bar{x}^2 - 4})^2$$

and

$$\bar{x}^2 - \tfrac{1}{4}(\bar{x} + \sqrt{\bar{x}^2 - 4})^2.$$

are monotone decreasing functions of $\bar{x}$.

An analogous result holds for the minimum when $n > 2$, but it does

not hold for the maximum. Thus, when n has any assigned value > 2, the x-coördinate

$$x' = \frac{1}{2^n}(\bar{x} - \sqrt{\bar{x}^2 - 4(n-1)})^n$$

of the minimum is a monotone decreasing function of $\bar{x}$, and approaches zero as a limit when $\bar{x}$ is increased indefinitely, but the mode does not, in general, approach the median $\bar{x}^n$ as a limit when n is increased. However, the ratio

$$\frac{1}{2^n \bar{x}^n}(\bar{x} + \sqrt{\bar{x}^2 - 4(n-1)})^n$$

of the mode to the median approaches the limit 1 as $\bar{x}$ is increased indefinitely.

The rapidity of approach to the limiting values depends on the smallness of the ratio $4(n-1)/\bar{x}^2$. Hence, in order that the frequency curve may descend rapidly to a minimum in the neighborhood of the discontinuity at $x' = 0$, and in order that the mode shall be relatively near the median, it is necessary that $4(n-1)/\bar{x}^2$ shall be small. This condition is clearly necessary in order that the new frequency curve shall have roughly the appearance of a normal curve when we neglect the part of this curve which belongs to the interval from $x' = 0$ to the minimum.

Case II. $0 < n < 1$.

In this case, make $n = 1/m$, where $m > 1$. This case thus includes the distribution obtained by taking positive integral roots of a set of variates. We shall limit our considerations to the principal real values of the functions.

The equation (2) may be written

$$y' = \frac{mx'^{m-1}}{\sqrt{2\pi}} e^{\frac{-(x'^m - \bar{x})^2}{2}}. \tag{5}$$

When $n < 1$, it follows from (3) that the mode is greater than the median of the new distribution. There is a minimum at $x' = 0$ when m is an odd number > 1, and we have in this case a minimum given by the negative value of x' obtained from (3). If $4(n-1)/\bar{x}^2$ is small, the value of the function for $x' < 0$ is too nearly zero to distinguish the curve from the x-axis when drawn on a scale suitable for reproduction on an ordinary page. Further, if $4(n-1)/\bar{x}^2$ is small, the curve for $x' > 0$ may be described roughly as having the general appearance of a normal curve, but differing from the normal curve both in that it is somewhat skew, and in that $y' = 0$ at a finite point.

Case III. $n < 0$.

Let $n = -m$.

Then the frequency curve becomes

$$y' = \frac{1}{mx'^{\frac{m+1}{m}}\sqrt{2\pi}} e^{\frac{-\left(x'^{-\frac{1}{m}} - \bar{x}\right)^2}{2}}. \tag{6}$$

By giving y' the value zero when $x' = 0$, the curve becomes continuous at the origin.

The distribution has a modal value

$$x' = \frac{(\sqrt{\bar{x}^2 + 4(m+1)} - \bar{x})^m}{2^m(m+1)^m}.$$

This mode is less than the median $1/\bar{x}^m$.

From the three cases examined relative to values of n, we may now state the theorem that *the nth powers of a set of normally distributed positive variates give a distribution whose modal value is greater or less than its median according as the value of n is or is not between* 0 *and* 1.

The simplicity of the examination of the frequency distributions obtained from a normal distribution by the transformation $\bar{x} = x^n$ arises from the fact that the equation

$$\frac{dy}{dx} = 0$$

has a quadratic factor in the variable $x = x'^{\frac{1}{n}}$ for which we solved to determine maxima and minima.

The occurrence of this quadratic factor suggests the problem of finding other functions

$$x' = f(x) \tag{7}$$

which would lead to a quadratic equation in x, and in more special cases to a linear equation in x, for finding maxima and minima of the frequency distribution obtained by the transformation (7).

Assume that (7) may be solved for x giving a single-valued function

$$x = \varphi(x'). \tag{8}$$

Then the frequency curve of x'-variates is

$$y' = \frac{1}{\sqrt{2\pi}} e^{\frac{-[\varphi(x') - \bar{x}]^2}{2}} \frac{dx}{dx'},$$

and

$$\frac{dy'}{dx'} = \frac{1}{\sqrt{2\pi}} e^{\frac{-[\varphi(x') - \bar{x}]^2}{2}} \left\{ \frac{d^2x}{dx'^2} - \left(\frac{dx}{dx'}\right)^2 (x - \bar{x}) \right\} = 0 \tag{9}$$

gives the maxima and minima.

We may now seek the function that will make the equation

$$\frac{d^2x}{dx'^2} - \left(\frac{dx}{dx'}\right)^2 (x - \bar{x}) = 0 \tag{10}$$

have a quadratic factor in x.

Such a quadratic factor would exist, in general, if $\left(\frac{dx}{dx'}\right)^2 \Big/ \frac{d^2x}{dx'^2}$ were a linear function of x, say $cx + c_1$.

That is,

$$cx\frac{d^2x}{dx'^2} + c_1\frac{d^2x}{dx'^2} = \left(\frac{dx}{dx'}\right)^2. \tag{11}$$

Let $p = dx/dx'$. Then (11) becomes

$$cxp\frac{dp}{dx} + c_1p\frac{dp}{dx} = p^2,$$

and apart from the trivial solution $x = a$ constant, we have the solutions

$$x' = c_2\left(x + \frac{c_1}{c}\right)^{1-\frac{1}{c}} = c_2[x + c_1(1 - n)]^n, \tag{12}$$

and

$$x' = c_3 \log\left(x + \frac{c_1}{c}\right). \tag{13}$$

Thus we find that the logarithms of the variates as well as their powers are distributed in accord with frequency curves whose maxima and minima are easily obtained because of the quadratic factor in (10) when $x' = \log x$. The frequency distribution for the transformation $x' = \log x$ is similar to that of the case $0 < n < 1$ discussed above in that the mode is greater than the median.

When $\left(\frac{dx}{dx'}\right)^2 \Big/ \frac{d^2x}{dx'^2} = c_1$, where c_1 is a constant, the equation (10) has, in general, a linear factor, and

$$x' = c_4e^{-\frac{x}{c_1}} + c_5. \tag{14}$$

Thus we find that a simple exponential transformation of variates leads to a linear factor in equation (10).

Another transformation that would, in general, lead to a quadratic factor in (10) is given by making

$$\frac{\dfrac{d^2x}{dx'^2}}{\left(\dfrac{dx}{dx'}\right)^2} = Ax^2 + Bx + C.$$

From this equation

$$x' = c_6 \int \frac{dx}{e^{\frac{Ax^3}{3} + \frac{Bx^2}{2} + Cx}},$$

which could hardly be regarded as a simple transformation in which we are likely to be interested unless $A = B = 0$, but this special case gives simply the transformation (14).

The University of Iowa,
Iowa City, Iowa.

THE ASSOCIATED POINT OF SEVEN POINTS IN SPACE.*

By H. S. White.

From seven points in space an eighth point can be constructed to complete what Hesse and later geometricians have called a set of *associated* points. Unless their relative situation is in some way specialized, the construction is unique; and from any seven of the complete set the eighth is determined by the same method. That is, the eight points are symmetrically related. The most interesting properties of the set relate to surfaces of the second order, and it was natural that Hesse and many writers after him should employ quadrics in demonstrating even the uniqueness of the eighth point according to their several modes of construction. But the construction itself is linear,—by means of lines and planes exclusively. Accordingly the demonstration of uniqueness and symmetry does not actually require the use of quadrics.

It is proposed here to follow Hesse's† construction, to obtain an explicit equation for the eighth point as a covariant of the first seven in the set of associated points, and to prove from the algebraic forms its uniqueness and the symmetry of the set. Particularly interesting are equations 15 and 16.

1. **Geometric construction.** The first step in construction is to select one of the seven given points as a center of projection, or a first Brianchon point; and to adopt some definite order of sequence for the other six, regarding them as the vertices of a simple gauche hexagon. We shall use the numeral 7 for the first Brianchon point, and 123456 in cyclic order for the vertices of the skew hexagon. Next, draw three lines through point 7, each intersecting a pair of opposite sides of the hexagon. These are taken as diagonals of a first derived hexagon of Brianchon type, and their intersections with the sides of the given hexagon as vertices of this derived hexagon, inscribed in the first.

The original hexagon shall be called H, the first derived hexagon A, and a second derived hexagon B. Vertices of H are already named $1, 2, 3, \cdots, 6$; and its sides are properly indicated by $12, 23, \cdots, 56, 61$. Vertices of A may be denoted by a_{12}, a_{23}, etc., showing on what side of H they lie; while its sides are named α_1, α_2, etc., the side α_1 containing vertices a_{61} and a_{12}, α_2 connecting a_{12} and a_{23}, etc.

* Presented to the American Mathematical Society, New York, April 23, 1921.

† O. Hesse, "De curvis et superficiebus secundi ordinis," Crelle's Journal, vol. 20 (1840), pp. 285–308. For full references see Encyk. der math. Wissenschaften, vol. III2, pp. 248–9.

Construct in the third place six additional lines, which will be proved to form a second Brianchon hexagon inscribed in H. The first additional line, β_1, shall lie in the plane of sides 61 and 12 of H, and intersect the sides α_3 and α_5 of A. It meets also the side α_1, since it lies with α_1 in the plane 612. Similarly β_2 is to meet α_4 and α_6 and lie in the plane 123, and so forth.

It is to be proved that these six lines form a second Brianchon hexagon B_1 inscribed like the first in H. *The point where its three diagonals meet is the point* 8, *whose determination is the object of this construction.* Obviously points 7 and 8 are reciprocally related to the other six. It is to be proved also that point 8 is unchanged when points 7 and 6 exchange rôles; from this it follows that all eight points are symmetrically related.

2. **Formulæ for the 8th point.** The problem whose repeated solution yields the desired formula is this: given the equations of two points and a plane:

$$u_\alpha = 0, \qquad u_\beta = 0, \qquad a_r = 0,$$

to find the equation of the point where the plane meets the line joining the given points. It is of course

$$a_\alpha u_\beta - a_\beta u_\alpha = 0. \tag{1}$$

We may adopt certain abbreviations for formulæ. Point 1 shall be understood as having coördinates $x_1^1, x_2^1, x_3^1, x_4^1$. The equation $u_1x_1^1 + u_2x_2^1 + u_3x_3^1 + u_4x_4^1 = 0$ may be condensed to $ux^1 = 0$. The determinant of the coördinates of points 1, 2, 3, 4 is denoted by 1234, separated from other symbols by $+$, $-$, or $\cdot$ when necessary.

For the order of points on the original hexagon H, adopt first 123456, and take point 7 as Brianchon point of the first derived hexagon A. The side α_1 is to join points

a_{61} where line 61 meets plane 734, and
a_{12} " " 12 " " 745.

These points have the equations

$$\begin{aligned} 7346\cdot ux^1 - 7341\cdot ux^6 &= 0, \\ 7451\cdot ux^2 - 7452\cdot ux^1 &= 0. \end{aligned} \tag{2}$$

Cyclic permutation gives the four other vertices of A. Write explicitly equations for the points a_{23}, a_{31} on α_3. Both α_1 and α_3 are to meet the plane 456, and their join-line in that plane is named β_5. *We wish to prove that* β_5 *intersects* β_6 *on the line* 56; that is, that the six lines β_1, β_2, etc., form a closed gauche hexagon B inscribed in H.

From equations (2), on the model of (1), we can write next the equation of the point where line α_1 meets plane 456.

$$(7346\cdot 4561 - 7341\cdot 4566)(7451\cdot ux^2 - 7452\cdot ux^1) - (7451\cdot 4562 - 7452\cdot 4561)(7346\cdot ux^1 - 7341\cdot ux^6) = 0.$$

The italicized determinant vanishes; and by the use of identities this equation becomes

$$(3)\quad 7346\cdot 4561(7451\cdot ux^2 - 7452\cdot ux^1) - 7456\cdot 4512(7346\cdot ux^1 - 7341\cdot ux^6) = 0.$$

Similarly the equation of the point where line α_3 meets plane 456 is written thus:

$$(4)\quad 7614\cdot 4563\cdot(7562\cdot ux^3 - 7563\cdot ux^2) - 7564\cdot 2563\cdot(7614\cdot ux^3 - 7613\cdot ux^4) = 0.$$

Both these points being in the plane 456, it is possible to express their equations linearly in ux^4, ux^5, and ux^6. Reduced by the aid of the usual identities equations (3) and (4) become respectively:

$$(3a)\quad 2561\cdot 7346\cdot 7451\cdot ux^4 + 4261\cdot 7346\cdot 7451\cdot ux^5 + 4521\cdot 7345\cdot 7461\cdot ux^6 = 0.$$

$$(4a)\quad 2563\cdot 7615\cdot 7643\cdot ux^4 - 2364\cdot 7635\cdot 7614\cdot ux^5 + 2354\cdot 7635\cdot 7614\cdot vx^6 = 0.$$

Any point on their join-line, β_5, has its equation compounded linearly of (3a) and (4a); and the point where β_5 intersects the line 56 has an equation containing only ux^5 and ux^6, hence ux^4 is to vanish in the combination. The point on β_5 and 56 is given therefore by the equation

$$(5)\quad \left\{\begin{array}{l} 2563\cdot 7615\cdot 4261\cdot 7346\cdot 7451 \\ -\ 2561\cdot 7451\cdot 2364\cdot 7635\cdot 7614 \end{array}\right\} ux^5 + \left\{\begin{array}{l} 2563\cdot 7615\cdot 4521\cdot 7345\cdot 7461 \\ +\ 2561\cdot 7451\cdot 2354\cdot 7635\cdot 7614 \end{array}\right\} ux^6 = 0.$$

This equation can be written so as to exhibit better its invariance under a group of permutations.

$$(5a)\quad 7415\cdot ux^5 \left\{\begin{array}{l} 6325\cdot 6347\cdot 6124\cdot 6157 \\ -\ 6324\cdot 6357\cdot 6125\cdot 6147 \end{array}\right\} - 7416\cdot ux^6 \left\{\begin{array}{l} 5326\cdot 5347\cdot 5124\cdot 5167 \\ -\ 5324\cdot 5367\cdot 5126\cdot 5147 \end{array}\right\} = 0.$$

Attend to the coefficients in braces. Each is of the form that we may call a *Pascalian*. Its vanishing is the condition that six points shall be pro-

jected from the seventh by rays of a cone of the second order. In the coefficient of ux^5, the point 6 is the center of projection; in that of ux^6, the point 5. It is known that when the center of projection is left unchanged, permutation of the other six points leaves a Pascalian invariant except as to sign. Odd permutations change the sign, even permutations do not. Accordingly *the point where line β_6 meets the side 56 is this same point on β_5.* For the points 4, 1 which appear in the factors 7415 and 7416 are adjacent to vertices 5 and 6 in the chosen order 123456; and have the same relation after the order is reversed: 432165. Also the two Pascalian coefficients are merely exchanged. Therefore the point may be designated indifferently by b_{56} or b_{65}, if it is remembered that 4 and 1 are adjacent to the pair of vertices 6 and 5.

A diagonal of hexagon B is completely described, without auxiliary memoranda, by naming the two opposite sides of hexagon H on which lie the two opposite vertices of B, e.g., $\beta(56, 23)$. It is unnecessary to mention the point 7 as different in function from 4 and 1, since the points 7, 4, 1 can be permuted among themselves without altering the position of the points b_{56} and b_{23}. For the Pascalians in equation (5*a*) are invariant under such permutation, and the external factors 7415, 7416 change sign simultaneously. Hence the notation $\beta(56, 23)$ or $\beta(65, 23)$, etc., cannot become ambiguous.

If this second derived hexagon, B, is of the Brianchon type, its three diagonals intersect. If that intersection is uniquely determined by the seven given points, *all diagonals of all second derived hexagons must intersect.* We shall prove that this is the case, and the uniqueness of the eighth point is thereby proved.

Equation (5*a*) shall be used as a model. To avoid any possible ambiguity we append all seven points in a definite order as index of a Pascalian; thus:

$$6325\cdot 6347\cdot 6124\cdot 6157 - 6324\cdot 6357\cdot 6125\cdot 6147 \equiv P_{6,\ 312547} \equiv -P_{6,\ 132547}.$$

Then equation (5*a*) can be rewritten:

$$\frac{7415\cdot ux^5}{P_{5,\ 312647}} - \frac{7416\cdot ux^6}{P_{6,\ 312547}} = 0. \tag{5b}$$

On this model, write equations of the points b_{12} and b_{34} as opposite vertices of a second derived hexagon.

$$\frac{7561\cdot ux^1}{P_{1,\ 234567}} - \frac{7562\cdot ux^2}{P_{2,\ 134567}} = 0. \tag{6}$$

$$\frac{7563\cdot ux^3}{P_{3,\ 124567}} - \frac{7564\cdot ux^4}{P_{4,\ 123567}} = 0. \tag{7}$$

If we add these equations, the point represented is certainly on the diagonal $\beta(12, 34)$. For it determines with 3 and 4 a plane containing the point b_{12}, and with 1 and 2 a plane containing the point b_{34}. Call this point $b(12, 34)$. Its equation is this:

$$(8)\qquad \frac{7561\cdot ux^1}{P_{1,\,234567}} - \frac{7562\cdot ux^2}{P_{2,\,134567}} + \frac{7563\cdot ux^3}{P_{3,\,124567}} - \frac{7564\cdot ux^4}{P_{4,\,123567}} = 0.$$

This point coincides with $b(12, 35)$, whose equation is the following:

$$(9)\qquad \frac{7461\cdot ux^1}{P_{1,\,235467}} - \frac{7462\cdot ux^2}{P_{2,\,135467}} + \frac{7463\cdot ux^3}{P_{3,\,125467}} - \frac{7465\cdot ux^5}{P_{5,\,123467}} = 0.$$

The identity of the two points is seen upon applying to (9) the relation

$$(10)\qquad 1234\cdot ux^5 \equiv 1235\cdot ux^4 + 1254\cdot ux^3 + 1534\cdot ux^2 + 5234\cdot ux^1$$

and consolidating the result by the aid of relations among three Pascalians like the following:

$$(11)\qquad 2345\cdot 2367\cdot P_{1,\,234567} - 1345\cdot 1367\cdot P_{2,\,134567} \equiv 1245\cdot 1267\cdot P_{3,\,214567}.$$

For we have after the first substitution:

$$0 = \left(\frac{7461\cdot 1234}{P_{1,\,235467}} - \frac{7465\cdot 5234}{P_{5,\,123467}}\right)ux^1 - \left(\frac{7462\cdot 1234}{P_{2,\,135467}} + \frac{7465\cdot 1534}{P_{5,\,123467}}\right)ux^2 + \left(\frac{7463\cdot 1234}{P_{3,\,125467}} - \frac{7465\cdot 1254}{P_{5,\,123467}}\right)ux^3 - \frac{7465\cdot 1235}{P_{5,\,123467}}\cdot ux^4,$$

and this becomes, after three applications of formulæ like (11),

$$(12)\qquad 0 = \frac{7165\cdot 5231\cdot P_{4,\,235167}}{P_{1,\,235467}\cdot P_{5,\,123467}}\cdot ux^1 - \frac{7265\cdot 1532\cdot P_{4,\,123567}}{P_{2,\,135467}\cdot P_{5,\,123467}}\cdot ux^2 + \frac{7365\cdot 1253\cdot P_{4,\,123567}}{P_{3,\,125467}\cdot P_{5,\,123467}}\cdot ux^3 - \frac{7465\cdot 1235}{P_{5,\,123467}}\cdot ux^4.$$

After removal of three factors, this is precisely equation (8). Therefore, as asserted above, the index of the diagonal can be changed by substitution of any one point, without disturbing the incidence of the line and the point $b(12, 34)$ given by equation (8). But in that manner in succession *all possible diagonals of second derived hexagons can be reached; therefore all contain this same 8th point,* whose uniqueness is thus proved.

In the construction of hexagons A and B is seen the reciprocal relation of their Brianchon points, the 7th and 8th of the associated set. The 7th is exchangeable with any of the first six; e.g., in equation (8) it is permutable

with the 6th or 5th, quite obviously, hence also with the others. *Therefore all eight associated points are symmetrically related.*

3. **The extraneous factor.** Equation (8) represents a class of 35 equations, all equivalent geometrically since all represent equally the same eighth point of the set. In conciseness and symmetry it is not likely that this equation can be surpassed. *It contains, however, when cleared of fractions, the extraneous factor* 1234. For from the reduction in equation (12) we see that

$$1234 \cdot H_{(12,\ 35)} \equiv -\ 1235 \cdot H_{(12,\ 34)} \tag{13}$$

where $H_{(12,\ 34)} = 0$ is the left side of equation (8) cleared of fractions. Dividing out this factor 1234 would leave a form which does not change, save in sign (a skew contravariant), under any permutation of the first seven points:

$$\frac{H_{(12,\ 34)}}{1234} \equiv -\frac{H_{(12,\ 35)}}{1235}, \qquad \text{etc.} \tag{14}$$

From the structure of equation (8) it might be conjectured that we are dealing with a particular case of a form symmetric in two sets of coördinates, (u) and (v). Replace therefore the particular plane 756 by a parameter plane (v), and change signs of some terms by writing for index of P always some *cyclic* permutation of the order 1, 234567; extend the summation to all seven points.

$$\frac{vx^1 \cdot ux^1}{P_{1,\ 234567}} + \frac{vx^2 \cdot ux^2}{P_{2,\ 345671}} + \cdots + \frac{vx^7 \cdot ux^7}{P_{7,\ 123456}} = 0. \tag{15}$$

or briefly

$$\sum_{i=1}^{7} \left(\frac{ux^i \cdot vx^i}{P_i} \right) = 0 = S(u,\ v).$$

This includes as particular cases all 35 equivalent equations of type (8). It is, for every plane (v), the equation of the same point (x^8), since it may be compounded linearly from any four equations of type (8) which have not a common index-point, e.g., those which correspond to the four faces of a tetrahedron 4567. Further, this point $S(v,\ u) = 0$ is the polar of plane (v) with respect to the quadric envelope

$$S(u,\ u) = 0. \tag{16}$$

But because all planes have the same polar point (x^8), this quadric locus is the bundle, counted double, of all planes through that point.

$$S(u,\ u) = [ux^8]^2.$$

This squared equation has the merit of lacking the extraneous factors which occur in type (8), and is indeed of degree 14 in the coördinates of each of the seven given points, so that according to *Sturm* (Math. Annalen, 1) *it is free from all extraneous factors.*

COMMON SOLUTIONS OF TWO SIMULTANEOUS PELL EQUATIONS.

By A. Arwin.

We shall in this brief paper discuss the two Pell equations

$$x^2 - 2y^2 = 1, \qquad y^2 - 3z^2 = 1 \tag{1}$$

relative to their common integral solutions. That $x = 3$, $y = 2$, $z = 1$ is such a solution we see immediately, and ask then: Do other integral solutions exist?

To answer this question we subtract one of our equations from the other, and get

$$x^2 - 3y^2 + 3z^2 = 0. \tag{2}$$

Every solution of this equation* may according to the general theory of numbers of the domain $K(\sqrt{-3})$ be written in the form

$$x = 3pq, \qquad y = \tfrac{1}{2}(3p^2 + q^2), \qquad z = \pm\tfrac{1}{2}(3p^2 - q^2), \tag{3}$$

where the double sign of z will be explained immediately. Introducing these values of y and z in (1) we get

$$q^4 - 12p^2q^2 + 9p^4 = -2. \tag{4}$$

The solutions of the second equation (1) are given by the equation

$$(y + z\sqrt{3}) = (2 + \sqrt{3})^r. \tag{5}$$

If $r \equiv 0 \pmod 3$ were possible, then $z \equiv 0 \pmod 3$, and hence from (3) $q = 0$. This, however, contradicts equation (4).

When $r = 3s_1 + 1$ we have

$$\begin{aligned} y + z\sqrt{3} &= (2 + \sqrt{3})^{3s_1+1}, \\ (2 - \sqrt{3})(y + z\sqrt{3}) &= (2 + \sqrt{3})^{3s_1}, \\ (2y - 3z) + \sqrt{3}(2z - y) &= (2 + \sqrt{3})^{3s_1}, \end{aligned} \tag{6}$$

or

$$2z - y \equiv -(z + y) \equiv 0 \pmod 3$$

from which follows that the sign $+$ must be used in the value for z in equation (3). When $r = 3s_2 - 1$ it follows in the same way that the sign $-$ must be used. Both of these cases satisfy equation (4).

* See for example Bachmann, P., Niedere Zahlentheorie, vol. II, p. 456.

Upon a closer examination of (4) we find in the first place that in the number domain $K(\sqrt{3})$ it may be factored as follows:

$$(q^2 - 6p^2 + 3\sqrt{3}\,p^2)(q^2 - 6p^2 - 3\sqrt{3}\,p^2) = (-2). \tag{7}$$

From

$$(q^2 - 6p^2 + 3\sqrt{3}\,p^2) = (q - \sqrt{3}\sqrt{2-\sqrt{3}}\,p)(q + \sqrt{3}\sqrt{2-\sqrt{3}}\,p) \tag{8}$$

follows then its final division into factors in the number domain $K(\sqrt{2-\sqrt{3}})$, which is a relative domain of $K(\sqrt{3})$ constructed on the unity $2 - \sqrt{3}$. This is a Galois domain which, on account of the relations

$$\sqrt{2+\sqrt{3}} - \sqrt{2-\sqrt{3}} = \sqrt{2}, \qquad \sqrt{2+\sqrt{3}} + \sqrt{2-\sqrt{3}} = \sqrt{6}, \tag{9}$$

is identical with the domain $K(\sqrt{2}, \sqrt{3})$ constructed from $K(\sqrt{2})$ and $K(\sqrt{3})$. Its defining equation may be written in the form

$$x^4 - 4x^2 + 1 = 0, \tag{10}$$

and a base is given in $1, \sqrt{3}, \sqrt{2-\sqrt{3}}, \sqrt{3}\sqrt{2-\sqrt{3}}$, which leads to the discriminant $d = 2^8 \cdot 3^2$ of the domain. To decide on the number of ideal classes in $K(\sqrt{2-\sqrt{3}})$ it is only necessary to examine the ideals whose norm* is $\leq \frac{4!}{4^4}\sqrt{d} = \frac{9}{2}$, i.e., the two prime numbers 2 and 3.

In $K(\sqrt{3})$ we have

$$(2) = (\sqrt{3}+1)(\sqrt{3}-1), \qquad (3) = (\sqrt{3})(\sqrt{3}), \tag{11}$$

where the parentheses indicate that there is a question of division into ideals. In $K(\sqrt{2-\sqrt{3}})$ we have

$$(\sqrt{3} - 1) = (1 - \sqrt{2-\sqrt{3}})(1 + \sqrt{2-\sqrt{3}})$$

from which we conclude that only one ideal class exists, which is the principal ideal class. From the ideal equation

$$(-2) = (1 - \sqrt{3}\sqrt{2-\sqrt{3}})(1 + \sqrt{3}\sqrt{2-\sqrt{3}}) \times (\sqrt{2-\sqrt{3}} - \sqrt{3})(\sqrt{2-\sqrt{3}} + \sqrt{3})$$

follows on account of (7) and (8) a number identity of one of the two types:

$$[1 - \sqrt{3}\sqrt{2-\sqrt{3}}][m_1 + m_2\sqrt{3} + (n_1 + n_2\sqrt{3})\sqrt{2-\sqrt{3}}] \equiv [q - p\sqrt{3}\sqrt{2-\sqrt{3}}]$$

or

$$[1 - \sqrt{3}\sqrt{2-\sqrt{3}}][m_1 + m_2\sqrt{3} + (n_1 + n_2\sqrt{3})\sqrt{2-\sqrt{3}}] \equiv [q\sqrt{2-\sqrt{3}} - p\sqrt{3}], \tag{12}$$

* Minkowski, H., Diophantische Approximationen, 1907, Theorem LIX, page 185.

where the expression in the second parenthesis represents a unit in $K(\sqrt{2-\sqrt{3}})$, and m_1, m_2, n_1, n_2, p and q are rational integers. From the general theory of number domains* we know of the existence in $K(\sqrt{2-\sqrt{3}})$ of three so-called fundamental units ϵ_1, ϵ_2, and ϵ_3, which have the property that every other unit E in $K(\sqrt{2-\sqrt{3}})$ may be written in the form:

$$E = \pm \epsilon_1^m \epsilon_2^n \epsilon_3^r, \tag{13}$$

m, n, and r being integers. In the case of $K(\sqrt{2-\sqrt{3}})$ we may for example take $\epsilon_1 = \sqrt{2+\sqrt{3}}$, $\epsilon_2 = 1+\sqrt{2}$, $\epsilon_3 = \sqrt{3}+\sqrt{2}$, and have then

$$\begin{aligned} E = \pm (\sqrt{2+\sqrt{3}})^{\eta_1}(1+\sqrt{2})^{\eta_2}(\sqrt{3}+\sqrt{2})^{\eta_3}(A+B\sqrt{3}) \\ \times (M+N\sqrt{2})(P+Q\sqrt{6}), \end{aligned} \tag{13'}$$

where the exponents η_1, η_2, η_3 independently of each other may take the values 0 or 1, and for which the equations

$$A^2 - 3B^2 = 1, \qquad M^2 - 2N^2 = 1, \qquad P^2 - 6Q^2 = 1, \tag{14}$$

are satisfied. We have the following relations:

$$\begin{aligned} &[\sqrt{2+\sqrt{3}}]^2 = 2+\sqrt{3} = [2+\sqrt{2-\sqrt{3}}][2-\sqrt{2-\sqrt{3}}], \\ &[\sqrt{3}+\sqrt{2}]^2 = 5+2\sqrt{6} = [2+\sqrt{2+\sqrt{3}}][2+\sqrt{2-\sqrt{3}}], \\ &\quad [1+\sqrt{2}]^2 = 3+2\sqrt{2} = [2+\sqrt{2+\sqrt{3}}][2-\sqrt{2-\sqrt{3}}], \\ &[1+\sqrt{2}][\sqrt{3}+\sqrt{2}] \\ &\qquad = 2+\sqrt{3}+[\sqrt{6}+\sqrt{2}] = \sqrt{2+\sqrt{3}}[2+\sqrt{2+\sqrt{3}}], \end{aligned} \tag{15}$$

which are not without importance.

From (12) the following system of equations is obtained:

$$\begin{aligned} 1\cdot m_1 + 0\cdot m_2 + 3n_1 - 6n_2 &= q(1) \text{ or } \; 0(2) \\ 0\cdot m_1 + 1\cdot m_2 - 2n_1 + 3n_2 &= 0 \quad \text{"} \; -p \\ 0\cdot m_1 - 3m_2 + 1\cdot n_1 + 0\cdot n_2 &= 0 \quad \text{"} \quad q \\ -1\cdot m_1 + 0\cdot m_2 + 0\cdot n_1 + 1\cdot n_2 &= -p \quad \text{"} \quad 0 \end{aligned} \tag{16}$$

that is to say in case (1)

$$m_1 = \frac{5q-3p}{2}, \qquad m_2 = 3\cdot\frac{q-p}{2}, \qquad n_1 = 9\cdot\frac{q-p}{2}, \qquad n_2 = 5\frac{q-p}{2},$$

or the two relations:

$$n_1 = 3m_2, \qquad 3n_2 = 5m_2, \tag{17}$$

* Hilbert, D., "Die Theorie der algebraischen Zahlkörper," Ber. der Deutsch. Math.-Verein., 1897, p. 214, Theorem 47.

and in case (2)

$$m_1 = 3\frac{q-3p}{2}, \qquad m_2 = \frac{q-5p}{2}, \qquad n_1 = 5\cdot\frac{q-3p}{2}, \qquad n_2 = 3\cdot\frac{q-3p}{2},$$

or

$$n_2 = m_1, \qquad 3n_1 = 5m_1. \tag{18}$$

Furthermore we have

$$[m_1^2 + 3m_2^2 - 2n_1^2 - 6n_2^2 + 6n_1n_2]^2 \\ - 3[2m_1m_2 - 4n_1n_2 + n_1^2 + 3n_2^2]^2 = 1.$$

By comparing the coefficients m_1, m_2, n_1, and n_2 in

$$m_1 + m_2\sqrt{3} + [n_1 + n_2\sqrt{3}]\sqrt{2-\sqrt{3}} \tag{19}$$

with those in (13′) we obtain the former expressed as functions of A, B, M, N, P, and Q. The different combinations η_1, η_2, $\eta_3 = 0$ or 1 must in this connection be treated separately. In the systems of equations (14) and (17) or (14) and (18) we have thus five equations with six unknown quantities. The purpose of this paper is now to show how a sixth independent relation may be found, by means of which an algebraic equation in one of the quantities A, B, etc., is obtained, and which equation we then shall have to examine only with reference to possible integral solutions. We shall in the following deal only with the case $\eta_1 = \eta_2 = \eta_3 = 0$. When we perform the substitution $(\sqrt{3}; -\sqrt{3})$ on E we find that on account of (9) $\sqrt{6}$ remains unchanged, while $\sqrt{2}$ changes into $\sqrt{-2}$, and a unit E_1 results which has the form:

$$E_1 = \pm(\sqrt{2-\sqrt{3}})^{\eta_1}(1-\sqrt{2})^{\eta_2}(-1)^{\eta_3}(\sqrt{3}+\sqrt{2})^{\eta_2} \\ \times (A - B\sqrt{3})(M - N\sqrt{2})(P + Q\sqrt{6}),$$

i.e.,

$$E\cdot E_1 = (-1)^{\eta_2+\eta_3}(5+2\sqrt{6})^{\eta_2}(P+Q\sqrt{6})^2. \tag{20}$$

If the same substitution is performed on (19), and if the values $m_1 = \alpha$, $m_2 = 3\beta$, $n_1 = 9\beta$, $n_2 = 5\beta$ are introduced, we get

$$E\cdot E_1 = (\alpha^2 - 21\beta^2) + \sqrt{6}(4\alpha\beta - 18\beta^2);$$

that is to say, when only the case $\eta_1 = \eta_2 = \eta_3 = 0$

$$P^2 + 6Q^2 = \alpha^2 - 21\beta^2, \qquad P^2 - 6Q^2 = 1$$

is considered, we get

$$2P^2 = \alpha^2 - 21\beta^2 + 1$$

or

$$6P^2 = 3m_1^2 - 7m_2^2 + 3. \tag{21'}$$

If the substitution $(\sqrt{2 - \sqrt{3}}; - \sqrt{2 - \sqrt{3}})$ is used, $\sqrt{6}$ changes into $- \sqrt{6}$, $\sqrt{2}$ into $- \sqrt{2}$, while $\sqrt{3}$ remains unchanged. Hence,

$$E \cdot E_2 = (-1)^{\eta_1+\eta_2}(2 + \sqrt{3})^{\eta_1}(A + B\sqrt{3})^2,$$
$$E \cdot E_2 = (\alpha^2 - 15\beta^2) + \sqrt{3}(6\alpha\beta - 24\beta^2),$$

or

$$A^2 + 3B^2 = \alpha^2 - 15\beta^2, \qquad A^2 - 3B^2 = 1,$$
$$2A^2 = \alpha^2 - 15\beta^2 + 1, \qquad 6A = 3m_1^2 - 5m_2^2 + 3. \tag{21''}$$

The two substitutions $(\sqrt{3}; - \sqrt{3})$ and $(\sqrt{2 - \sqrt{3}}; - \sqrt{2 - \sqrt{3}})$ used simultaneously give us

$$E \cdot E_3 = (-1)^{\eta_1+\eta_3}(3 + 2\sqrt{2})^{\eta_2}(M + N\sqrt{2})^2,$$
$$E \cdot E_3 = (\alpha^2 - 33\beta^2) + \sqrt{2}(6\alpha\beta - 36\beta^2),$$

or

$$6M = 3m_1^2 - 11m_2^2 + 3. \tag{21'''}$$

Eliminating m_1 and m_2 from the three equations (21′), (21″), and (21‴) we get

$$3P^2 - 2A^2 = M_1^2 \tag{22}$$

which for case (1), in which $\eta_1 = \eta_2 = \eta_3 = 0$, gives us the sixth independent equation. To show that this really is the case, we eliminate the four variables B, M, N, and Q, and obtain the two equations:

$$\begin{aligned} - 4A^8 - 48A^6P^2 + 32A^6 + 132A^4P^4 - 60A^4P^2 - 64A^4 - 72A^2P^6 \\ - 24A^2P^4 + 144A^2P^2 - 9P^8 + 54P^6 - 81P^4 = 0, \end{aligned} \tag{23'}$$

and

$$\begin{aligned} - 4A^8P^8 - 64A^8P^6 + 80A^8P^4 + 96A^8P^2 - 144A^8 + 8A^6P^8 \\ + 272A^6P^6 - 376A^6P^4 + 96A^6P^2 - 144A^6 + 32A^4P^8 - 28A^4P^6 \\ + 296A^4P^4 - 192A^4P^2 - 36A^4 - 36A^2P^8 - 828A^2P^6 + 972A^2P^4 \\ + 324A^2P^2 - 81P^8 + 486P^6 - 729P^4 = 0. \end{aligned} \tag{24'}$$

From (23′) we see then in the first place that no factor can be found which is independent of A. Furthermore, if (23′) were reducible, it must remain so for any arbitrary value of P^2, for example for $P^2 = - 1$. For this value of P^2 (23′) may be reduced to the form

$$A^8 - 20A^6 - 32A^4 + 24A^2 + 36 = 0, \tag{23''}$$

which by a simple discussion may be shown to be irreducible. For the same value $P^2 = - 1$ of P we obtain from (24′) the equation

$$25A^8 + 220A^6 - 128A^4 - 360A^2 + 324 = 0 \tag{24''}$$

from which it is seen that (23′) and (24′) really are distinct equations, and that the elimination from these of, for example, P^2 will lead to the desired

algebraic equation in A^2. This equation must then be discussed for possible integral solutions, which in the first place must satisfy equations (14). Finally we may easily verify that (23′) and (24′) are indeed satisfied by $A^2 = P^2 = 1$, which give us the already known solution $p = q = 1$.

In this way every combination $\eta_1, \eta_2, \eta_3 = 0$ or 1 must be tried in the two cases (1) and (2). Thus we find that our problem is completely solved by a finite number of purely algebraic operations. It is possible that a discussion of (22), (14), and (17) with reference only to divisibility would show that no other solution than the one mentioned could exist, and that thus in this special case the long process of elimination could be obviated. A similar method may be applied on equations of the type

$$ax^4 + 2bx^2y^2 + cy^4 = A,$$

where a, b, c, and A are given integers, whenever the ultimate relative domain is a Galois domain, as in the above example.

LUND,
SWEDEN,
October, 1920.

ON THE COMPLETE INDEPENDENCE OF HURWITZ'S POSTULATES FOR ABELIAN GROUPS AND FIELDS.*

By B. A. Bernstein.

In these Annals, in 1913, Hurwitz presented sets of postulates for abelian groups and fields—three for abelian groups (finite, denumerably infinite, and non-denumerably infinite) and three for corresponding fields.† The chief characteristics of each of these sets are the simplicity of the statements, the small number of postulates used, and the elegance of the systems establishing (ordinary) independence.‡ The object of this paper is to consider for these admirable sets of postulates the question of complete independence,§ which question Professor Hurwitz left untouched.

Hurwitz's postulates. Hurwitz's postulates are found among the following eight conditions on a class K and two binary operations $\oplus$, $\odot$.

(A_1) If a, b, c, $a \oplus b$, $c \oplus b$, and $a \oplus (c \oplus b)$ belong to K, then $(a \oplus b) \oplus c = a \oplus (c \oplus b)$.

(A_2) If a and b belong to K, then there is an element x of K such that $a \oplus x = b$.

(M_1) If a, b, c, $a \odot b$, $c \odot b$, and $a \odot (c \odot b)$ belong to K, then $(a \odot b) \odot c = a \odot (c \odot b)$.

(M_2) If a and b belong to K, and $a \oplus a \neq a$, there is an element x of K such that $a \odot x = b$.

(D) If a, b, c, $a \odot b$, $a \odot c$, $b \oplus c$, $(a \odot b) \oplus (a \odot c)$ belong to K, then $a \odot (b \oplus c) = (a \odot b) \oplus (a \odot c)$.

(N_n) K contains n (> 1) elements.

(N') K is countably infinite.

* Read before the San Francisco Section of the American Mathematical Society, October 22, 1921.

† W. A. Hurwitz, "Postulate-sets for abelian groups and fields," these Annals (2), vol. 15 (1913), p. 93. Compare his "Note on the definition of an abelian group," the Annals (2), vol. 8 (1907), p. 94.

‡ The postulates are based on sets of postulates for abelian groups and fields given by Huntington. See E. V. Huntington, "Definitions of a field by sets of independent postulates," Trans. Amer. Math. Soc., vol. 4 (1903), p. 31, and "Note on the definitions of abstract groups and fields by sets of independent postulates," Trans. Amer. Math. Soc., vol. 6 (1905), p. 181. While retaining the elegance of Huntington's postulates, Hurwitz reduces their number by one for abelian groups and by two for fields.

§ Professor E. H. Moore first proposed the problem of complete independence of a set of postulates. See his "Introduction to a form of general analysis," New Haven Mathematical Colloquium, Yale University Press, p. 82. On the significance of the question of complete independence of postulates see also E. V. Huntington, Bull. Amer. Math. Soc., vol. 23 (1917), p. 278, and J. S. Taylor, Bull. Amer. Math. Soc., vol. 26 (1920), p. 449, footnote.

(N'') K has the cardinal number of the continuum.

Let G_n, G', G'', F_n, F', F'' denote the sets taken from the above "matrix" as follows:

$$
\begin{aligned}
G_n&: (A_1),\ (A_2),\ (N_n),\\
G'&: (A_1),\ (A_2),\ (N'),\\
G''&: (A_1),\ (A_2),\ (N''),\\
F_n&: (A_1),\ (A_2),\ (M_1),\ (M_2),\ (D),\ (N_n),\\
F'&: (A_1),\ (A_2),\ (M_1),\ (M_2),\ (D),\ (N'),\\
F''&: (A_1),\ (A_2),\ (M_1),\ (M_2),\ (D),\ (N'').
\end{aligned}
$$

Hurwitz proves that G_n, G', G'' form sets of independent postulates for abelian groups having respectively n elements, a countable infinity of elements, and elements whose cardinal number is that of the continuum; and he proves that F_n,* F', F'' form sets of independent postulates for corresponding fields.

Complete independence. The question of complete independence of the postulate-sets is answered by the following

THEOREM. *Postulate-sets F', F'', G', G'', G_n $(n > 1)$ are each completely independent; postulate-set F_n is completely independent when, and only when, n exceeds 2 and is a power of a prime.*

To prove the complete independence of F' we take for systems having the characters $(\pm\pm\pm\pm\pm\pm+)$ systems 1–32 in Table A below. By taking for K in this table the class of *reals*, instead of the class of rationals, we obtain systems, 1′–32′, having the characters $(\pm\pm\pm\pm\pm-)$.

That set F'' is completely independent is seen from the fact that, with respect to F'', systems 1–32 have the characters $(\pm\pm\pm\pm\pm-)$, while systems 1′–32′ have the characters $(\pm\pm\pm\pm\pm+)$.

Since G' is included in F' and G'' in F'', postulate-sets G', G'' are each completely independent.

Proof-systems showing the complete independence of G_n are systems 4, 5, 6, 16,† together with the systems obtained from 4, 5, 6, 16* by replacing (1) the class of rationals by the class of n integers $0, 1, \cdots, n-1$ $(n > 1)$ and (2) the operation $a + b$ (in 4) by the operation $a + b \bmod n$.

In order to see that F_n is completely independent for every integer $n > 2$ and a power of a prime, we observe (1) that with respect to F_n systems 1–32 of Table A have the characters $(\pm\pm\pm\pm\pm\pm-)$; (2) that the Galois field of order $n = q^k$, q prime and $n > 2$, gives the character $(++++++)$; and (3) that systems 2–32 will have the remaining 31 of the 32 characters $(\pm\pm\pm\pm\pm+)$ if in these systems we replace (1) the

* When n is a power of a prime.

† As far as K, $\oplus$ are concerned.

class of rationals by the class of n integers $0, 1, \cdots, n-1$ ($n > 2$ and a power of a prime) and (2) the operation $a + b$ by $a + b$ mod n.

TABLE A.

SYSTEMS HAVING THE CHARACTERS $(\pm\ \pm\ \pm\ \pm\ \pm\ +)$ FOR F'.

No.	Character.	K.	$a \oplus b$.	$a \odot b$.
1	$(++++++)$	Rationals	$a+b$	ab
2	$(++++-+)$	"	$a+b$	$a+b$
3	$(+++-++)$	"	$a+b$	0
4	$(++-+++)$	"	$a+b$	b
5	$(+-++++)$	"	a	$a+b$
6	$(-+++++)$	"	b	$a+b$
7	$(+++--+)$	"	$a+b$	1
8	$(++-+-+)$	"	$a+b$	$b+1$
9	$(+-++-+)$	"	0	$a+b$
10	$(-+++-+)$	"	b except: $2 \oplus 0 = 1$ $2 \oplus 1 = 0$	0 except: $1 \odot 1 = 1$
11	$(++--++)$	"	$a+b$	b except: $1 \odot a = 0$
12	$(+-+-++)$	"	0	0
13	$(-++-++)$	"	$b+1$	y^*
14	$(+--+++)$	"	0	b
15	$(-+-+++)$	"	b	b
16	$(--++++)$	"	y^* except: $a \oplus a = a$ $0 \oplus 1 = 1$ $1 \oplus 0 = 0$	G
17	$(++---+)$	"	$a+b$	$a+1$
18	$(+-+--+)$	"	0	1
19	$(-++--+)$	"	$b+1$	1
20	$(+--+-+)$	"	0	$b+1$
21	$(-+-+-+)$	"	b except: $2 \oplus 0 = 1$ $2 \oplus 1 = 0$	0 except: $2 \odot 0 = 1$
22	$(--++-+)$	"	0 except: $1 \oplus 0 = 1$	$a+b$
23	$(+---++)$	"	0	0 except: $0 \odot 1 = 1$
24	$(-+--++)$	"	$b+1$	b except: $0 \odot a = y^*$
25	$(--+-++)$	"	1 except: $1 \oplus 0 = 0$	1
26	$(---+++)$	"	$a+1$	b
27	$(+----+)$	"	0	$a+1$
28	$(-+---+)$	"	$b+1$	1 except: $0 \odot 1 = 0$
29	$(--+--+)$	"	0 except: $0 \oplus 0 = 1$	0
30	$(----++)$	"	b except: $1 \oplus 1 = 0$	b except: $1 \odot 1 = 0$
31	$(---+-+)$	"	0 except: $0 \oplus 0 = 1$	$b+1$
32	$(-----+)$	"	$a+1$	$a+1$

Finally, F_n is *not* completely independent when n is other than a power of a prime, or when $n = 2$, because (1) there exists no field for n

* Not an element of K.

other than a power of a prime, and (2) there exists no system of character $(-+-+-+)$ when $n = 2$.* This completes the proof of our theorem.

If we only wish to prove the complete independence of sets F', F'', G', G'', systems 1°–32° of Table B below will be found more simple than systems 1–32 above.

TABLE B.

SYSTEMS HAVING THE CHARACTERS $(\pm\pm\pm\pm\pm+)$ FOR F'.

No.	Character.	K.	$a \oplus b$.	$a \odot b$.
1°	$(++++++)$	Rationals†	$a + b$	ab
2°	$(++++-+)$	"	$a + b$	$a + b$
3°	$(+++-++)$	"	$a + b$	0
4°	$(++-+++)$	"	$a + b$	b/a
5°	$(+-++++)$	"	0	ab
6°	$(-+++++)$	"	$a - b$	ab
7°	$(+++--+)$	"	$a + b$	1
8°	$(++-+-+)$	"	$a + b$	$a - b$
9°	$(+-++-+)$	"	0	$a + b$
10°	$(-+++-+)$	"	$a - b$	$a + b$
11°	$(++--++)$	"	$a + b$	$b/(a - 1)$
12°	$(+-+-++)$	"	0	0
13°	$(-++-++)$	"	$a - b$	0
14°	$(+--+++)$	"	0	b
15°	$(-+-+++)$	"	b	b
16°	$(--++++)$	"	$a/2$	ab
17°	$(++---+)$	"	$a + b$	a/b
18°	$(+-+--+)$	"	0	1
19°	$(-++--+)$	"	$a - b$	1
20°	$(+--+-+)$	"	0	$a - b$
21°	$(-+-+-+)$	"	$a - b$	$a - b$
22°	$(--++-+)$	"	a/b	$a + b$
23°	$(+---++)$	"	0	$(a - 1)b$
24°	$(-+--++)$	"	$a - b$	$b/(a - 1)$
25°	$(--+-++)$	"	a/b	1
26°	$(---+++)$	"	$a/2$	b/a
27°	$(+----+)$	"	0	$a + 1$
28°	$(-+---+)$	"	$a - b$	a/b
29°	$(--+--+)$	"	a/b	ab
30°	$(----++)$	"	$a/2$	$(a - 1)b$
31°	$(---+-+)$	"	a/b	$(a - 1)b$
32°	$(-----+)$	"	a/b	a/b

UNIVERSITY OF CALIFORNIA,
October, 1921.

* If 0, 1 be the two elements of K, the only choice we have for $a \oplus b$ so that postulate (A_2) be satisfied is:

$$\overset{(1)}{\begin{array}{c|cc} \oplus & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 0 & 1 \end{array}}, \qquad \overset{(2)}{\begin{array}{c|cc} \oplus & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array}}, \qquad \overset{(3)}{\begin{array}{c|cc} \oplus & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 0 & 1 \end{array}}, \qquad \overset{(4)}{\begin{array}{c|cc} \oplus & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 1 & 0 \end{array}},$$

i.e., respectively

$$a \oplus b = b, \qquad a + b \bmod 2, \qquad a + b + 1 \bmod 2, \qquad b + 1 \bmod 2.$$

Of these, system (4) is the only one which contradicts both (A_1) and (D). System (4) is likewise the only possibility for $a \odot b$ in order that both (M_1) and (D) be contradicted. But if (4) be taken for both $a \oplus b$ and $a \odot b$, postulate (D) will be satisfied.

† *All* the rationals—positive, negative, and zero.

ON POWER SERIES WITH POSITIVE REAL PART IN THE UNIT CIRCLE.*

By T. H. Gronwall.

1. **Introduction.** Let $\varphi(z)$ be a power series convergent for $|z| < 1$ and such that $\Re\varphi(z) \geqq 0$ in the unit circle. Since the real part of a function holomorphic in the unit circle cannot have a minimum inside the circle without being a constant, it follows that $\Re\varphi(z) > 0$ for $|z| < 1$ unless $\varphi(z)$ is a purely imaginary constant. Disregarding this trivial case, it is seen that multiplying $\varphi(z)$ by a positive constant, we may make $\Re\varphi(0) = \frac{1}{2}$, and subtracting a purely imaginary constant, we may therefore assume $\varphi(z)$ to be of the form

$$\varphi(z) = \tfrac{1}{2} + \sum_{\nu=1}^{\infty} a_\nu z^\nu. \tag{1}$$

The following question now arises: what conditions must the constants $a_1, a_2, \cdots, a_n$ satisfy in order that there shall exist a $\varphi(z)$ of the form (1), convergent for $|z| < 1$, having the given constants as its first n coefficients, and such that $\Re\varphi(z) > 0$ for $|z| < 1$? It has been shown by Carathéodory,† by methods belonging to Minkowski's theory of convex solids, that all $a_1, a_2, \cdots, a_n$ with the required property are interior to or on the boundary of a certain convex solid K_n‡ and may be uniquely represented in parametric form by

$$a_\nu = \lambda_1 e^{-\nu\alpha_1 i} + \lambda_2 e^{-\nu\alpha_2 i} + \cdots + \lambda_n e^{-\nu\alpha_n i} \quad (\nu = 1, 2, \cdots, n) \tag{2}$$

where the α's lie between 0 and 2π (incl.), the λ's are positive or zero, and

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n < 1$$

for points interior to K_n, but

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = 1$$

when the point $a_1, a_2, \cdots, a_n$ is on the boundary of K_n. In the latter

* Read before the American Mathematical Society, September 7, 1921.

† "Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen," Math. Annalen, vol. 64 (1907), pp. 95–115, and "Über den Variabilitätsbereich der Fourier'schen Konstanten von positiven harmonischen Funktionen," Rendiconti del Circolo Matematico di Palermo, vol. 32 (1911), pp. 193–217.

‡ That is, writing $a_\nu = x_\nu + ix_{n+\nu}$ ($\nu = 1, 2, \cdots, n$), the points of rectangular coördinates $x_1, \cdots, x_{2n}$ form a convex solid in Euclidean $2n$-space.

case, $\varphi(z)$ is uniquely determined by the coefficients $a_1, a_2, \cdots, a_n$, and has the form

$$(3)\qquad \varphi(z) = \frac{1}{2}\lambda_1\frac{e^{\alpha_1 t}+z}{e^{\alpha_1 t}-z} + \frac{1}{2}\lambda_2\frac{e^{\alpha_2 t}+z}{e^{\alpha_2 t}-z} + \cdots + \frac{1}{2}\lambda_n\frac{e^{\alpha_n t}+z}{e^{\alpha_n t}-z}$$

(where $\lambda_1 \geqq 0, \lambda_2 \geqq 0, \cdots, \lambda_n \geqq 0$ and $\lambda_1 + \lambda_2 + \cdots + \lambda_n = 1$).

The convex solid K_n may also be defined by algebraic inequalities involving $a_1, a_2, \cdots, a_n$ and their conjugates $\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_n$, as was shown by Toeplitz* and Fischer† through the consideration of certain definite Hermitian forms. Writing $D_0 = 1$ and

$$(4)\qquad D_m = D_m(a_1, a_2, \cdots, a_m) = \begin{vmatrix} 1 & a_1 & a_2 & \cdots & a_m \\ \bar{a}_1 & 1 & a_1 & \cdots & a_{m-1} \\ \bar{a}_2 & \bar{a}_1 & 1 & \cdots & a_{m-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \bar{a}_m & \bar{a}_{m-1} & \bar{a}_{m-2} & \cdots & 1 \end{vmatrix}$$

for $m = 1, 2, \cdots, n$, the necessary and sufficient condition that $a_1, a_2, \cdots, a_n$ shall be interior to K_n is

$$(5)\qquad D_0 > 0, \qquad D_1 > 0, \qquad D_2 > 0, \qquad \cdots, \qquad D_n > 0,$$

while for a point on the boundary of K_n it is necessary and sufficient that there shall exist a k, where $1 \leqq k \leqq n$, such that

$$(6)\quad D_0 > 0, \quad D_1 > 0, \quad \cdots, \quad D_{k-1} > 0, \quad D_k = D_{k+1} = \cdots = D_n = 0.$$

The preceding results were also obtained independently by F. Riesz.‡

It is the purpose of the present paper to prove all these results by the most elementary function theoretic means, the method of treatment resembling closely that of a preceding paper by the writer§ dealing with a similar problem first solved by Carathéodory and Fejér. The central part of the argument consists in the combination of the process of complete induction with Schwarz' lemma, and thus furnishes a new and not uninteresting example of the fundamental importance of the latter in the theory of functions of a complex variable.

2. **The point set K_n and its correspondence with K_{n-1}.** We begin by recalling some familiar definitions. A sequence of n complex numbers $a_1, a_2, \cdots, a_n$ is called a point (all a's are assumed to be finite). The

* "Über die Fourier'sche Entwickelung positiver Funktionen," Rendiconti del Circolo Matematico di Palermo, vol. 32 (1911), pp. 191–192.

† "Über das Carathéodory'sche Problem, Potenzreihen mit positivem reellen Teil betreffend," ibid., pp. 240–256.

‡ "Sur certains systèmes singuliers d'équations intégrales," Annales de l'Ecole Normale, ser. 3, vol. 28 (1911), pp. 33–62.

§ "On the maximum modulus of an analytic function," these ANNALS, ser. 2, vol. 16 (1914), pp. 77–81.

neighborhood ϵ of a point $a_1, a_2, \cdots, a_n$ consists of all points $a_1', a_2', \cdots, a_n'$ such that

$$|a_1' - a_1| < \epsilon, \qquad |a_2' - a_2| < \epsilon, \qquad \cdots, \qquad |a_n' - a_n| < \epsilon.$$

Consider any point set P. A boundary point of P is any point such that every neighborhood ϵ of this point contains a point belonging to P and also a point not belonging to P; the boundary point itself may or may not belong to P. To every point not on the boundary of P there consequently exists an ϵ such that the neighborhood ϵ of this point consists either of points all belonging to P or of points none of which belongs to P. In the former case, the point is said to be interior to P, and in the latter case, exterior to P. It follows that an interior point belongs to P, while an exterior point does not.

We now define K_n as the set of all points $a_1, a_2, \cdots, a_n$ such that there exists a power series $\varphi(z) = \frac{1}{2} + a_1 z + a_2 z^2 + \cdots + a_n z^n + \cdots$ convergent and with positive real part for $|z| < 1$. Any such $\varphi(z)$ is said to be associated with the point $a_1, a_2, \cdots, a_n$.

Then K_n *contains interior points,* for assuming $|a_1| < 1/2n$, $|a_2| < 1/2n$, $\cdots$, $|a_n| < 1/2n$, the point $a_1, a_2, \cdots, a_n$ belongs to K_n, since the polynomial $\varphi(z) = \frac{1}{2} + a_1 z + a_2 z^2 + \cdots + a_n z^n$ has the required properties on account of $\Re(a_\nu z^\nu) \geqq -|a_\nu z^\nu| > -1/2n$ for $|z| < 1$ and $\nu = 1, 2, \cdots, n$. Consequently, any point $a_1, a_2, \cdots, a_n$ where $|a_1| < 1/4n$, $|a_2| < 1/4n$, $\cdots$, $|a_n| < 1/4n$ has a neighborhood $1/4n$ containing only points of K_n, which proves our statement.

We shall now perform a sequence of transformations which will finally lead to a correspondence between K_n and K_{n-1}. First, consider a $\varphi(z)$ associated with a point of K_n; then $\varphi(z) + \frac{1}{2}$ does not vanish for $|z| < 1$, its real part being greater than $\frac{1}{2}$, and consequently

$$f(z) = \frac{\varphi(z) - \frac{1}{2}}{\varphi(z) + \frac{1}{2}} = a_1 z + \cdots \tag{7}$$

is holomorphic for $|z| < 1$; moreover, the identity

$$\begin{aligned} 1 - |f|^2 &= 1 - f\bar{f} \\ &= 1 - \frac{\varphi - \frac{1}{2}}{\varphi + \frac{1}{2}} \cdot \frac{\bar{\varphi} - \frac{1}{2}}{\bar{\varphi} + \frac{1}{2}} = \frac{\varphi + \bar{\varphi}}{(\varphi + \frac{1}{2})(\bar{\varphi} + \frac{1}{2})} = \frac{2\Re\varphi}{|\varphi + \frac{1}{2}|^2} \end{aligned} \tag{8}$$

shows that $|f(z)| < 1$ for $|z| < 1$ since $\Re\varphi(z) > 0$. Conversely, (7) gives

$$\varphi(z) = \frac{1}{2}\frac{1 + f(z)}{1 - f(z)}, \tag{9}$$

and from (8) and (9) we obtain

$$2\Re\varphi = \frac{1 - |f|^2}{|1 - f|^2}. \tag{10}$$

From (9) and (10) it follows that, for any $f(z)$ holomorphic and less than unity in absolute value for $|z| < 1$, (9) defines a $\varphi(z)$ holomorphic and with positive real part for $|z| < 1$, and if $f(0) = 0$, so that $f(z) = a_1 z + \cdots$, then $\varphi(z) = \frac{1}{2} + a_1 z + \cdots$.

Now let $f(z) = a_1 z + \cdots$ be any function vanishing at the origin, and holomorphic and less than unity in absolute value for $|z| < 1$. Writing

$$g(z) = \frac{1}{z} f(z) = a_1 + \cdots \tag{11}$$

it follows from Schwarz' lemma that

$$|g(z)| \leqq 1 \qquad \text{for} \qquad |z| < 1, \tag{12}$$

and, if $|g(z)| = 1$ for a value of z inside the unit circle, then $g(z)$ is constant $= a_1$, where $|a_1| = 1$. Conversely, any function $g(z)$ holomorphic and less than or equal to unity in absolute value for $|z| < 1$ defines an $f(z) = zg(z)$ holomorphic for $|z| < 1$, and $|f(z)| < 1$ for $|z| < 1$. Thus we always have

$$|a_1| \leqq 1. \tag{13}$$

It now follows that the point set K_1 is defined by (13), so that its boundary points, given by $|a_1| = 1$, belong to K_1, and that with any $a_1 = e^{-\alpha i}$ $(0 \leqq \alpha < 2\pi)$ on the boundary of K_1 there is associated one and only one $\varphi(z)$, namely

$$\varphi(z) = \frac{1}{2}\frac{e^{\alpha i} + z}{e^{\alpha i} - z}. \tag{14}$$

In fact, consider any $\varphi(z) = \frac{1}{2} + a_1 z + \cdots$ holomorphic and of positive real part for $|z| < 1$; then (7) and (11) define a $g(z) = a_1 + \cdots$ satisfying (12), and making $z = 0$ in (12), we obtain (13). Conversely, taking any a_1 such that $|a_1| \leqq 1$, and making $g(z) = a_1$, (12) is satisfied, and (11) and (9) give

$$\varphi(z) = \frac{1}{2}\frac{1 + a_1 z}{1 - a_1 z} \tag{15}$$

as one of the functions satisfying all the conditions imposed on $\varphi(z)$. Moreover, when $|a_1| = 1$, $g(z)$ is uniquely determined and equals a_1, so that (15) is the only φ-function possible, and writing $a_1 = e^{-\alpha i}$, we obtain (14).

Now assume $|a_1| < 1$, then it follows from what precedes that (12) takes the form

$$|g(z)| < 1 \qquad \text{for} \qquad |z| < 1. \tag{12'}$$

Writing

$$f_1(z) = \frac{g(z) - a_1}{1 - \bar{a}_1 g(z)}, \tag{16}$$

we have $|\bar{a}_1 g(z)| < 1$ for $|z| < 1$, so that $f_1(z)$ is holomorphic for $|z| < 1$, and $f_1(0) = 0$; moreover, the identity

$$1 - |f_1|^2 = 1 - \frac{g - a_1}{1 - \bar{a}_1 g} \cdot \frac{\bar{g} - \bar{a}_1}{1 - a_1 \bar{g}} = \frac{(1 - |a_1|^2)(1 - |g|^2)}{|1 - \bar{a}_1 g|^2} \tag{17}$$

shows that

$$|f_1(z)| < 1 \qquad \text{for} \qquad |z| < 1. \tag{18}$$

Conversely, to any $f_1(z)$ holomorphic and less than unity in absolute value for $|z| < 1$, and vanishing at the origin, there corresponds a $g(z)$ obtained from (16)

$$g(z) = \frac{f_1(z) + a_1}{1 + \bar{a}_1 f_1(z)}; \tag{19}$$

this $g(z)$ is holomorphic for $|z| < 1$, $g(0) = a_1$, and $|g(z)| < 1$ for $|z| < 1$, as is seen by interchanging g and f_1 and replacing a_1 by $-a_1$ in (17). Finally, we write

$$\varphi_1(z) = \frac{1}{2}\frac{1 + f_1(z)}{1 - f_1(z)}, \qquad f_1(z) = \frac{\varphi_1(z) - \frac{1}{2}}{\varphi_1(z) + \frac{1}{2}}; \tag{20}$$

it follows from what has been said above in regard to $\varphi(z)$ and $f(z)$ that when $f_1 = b_1 z + \cdots$ is holomorphic and less than unity in absolute value for $|z| < 1$, then $\varphi_1(z) = \frac{1}{2} + b_1 z + \cdots$ is holomorphic and has its real part positive for $|z| < 1$, and vice versa.

We have thus proved that to every $\varphi(z) = \frac{1}{2} + a_1 z + \cdots + a_n z^n + \cdots$, holomorphic and with positive real part for $|z| < 1$, and such that $|a_1| < 1$, there corresponds uniquely, by means of (7), (11), (16) and (20), a $\varphi_1(z) = \frac{1}{2} + b_1 z + \cdots + b_{n-1} z^{n-1} + \cdots$ holomorphic and with positive real part for $|z| < 1$. Conversely, to a given $\varphi_1(z) = \frac{1}{2} + b_1 z + \cdots + b_{n-1} z^{n-1} + \cdots$ holomorphic and with positive real part for $|z| < 1$, and a given a_1 where $|a_1| < 1$, there corresponds uniquely a $\varphi(z) = \frac{1}{2} + a_1 z + \cdots + a_n z^n + \cdots$ holomorphic and of positive real part for $|z| < 1$. It will be necessary for the following to establish the general form of the relation between the coefficients a and b. From (19) and (20) we find

$$\begin{aligned} g(z) &= \frac{(1 + a_1)\varphi_1(z) - \frac{1}{2}(1 - a_1)}{(1 + \bar{a}_1)\varphi_1(z) + \frac{1}{2}(1 - \bar{a}_1)} \\ &= \frac{a_1 + (1 + a_1) b_1 z + \cdots + (1 + a_1) b_{n-1} z^{n-1} + \cdots}{1 + (1 + \bar{a}_1) b_1 z + \cdots + (1 + \bar{a}_1) b_{n-1} z^{n-1} + \cdots} \\ &= a_1 + g_1 z + g_2 z^2 + \cdots + g_{n-1} z^{n-1} + \cdots, \end{aligned}$$

where

$$g_1 = (1 - a_1\bar{a}_1)b_1,$$
$$g_\nu = (1 - a_1\bar{a}_1)b_\nu + G_\nu(a_1, \bar{a}_1, b_1, b_2, \cdots, b_{\nu-1}),$$

for $\nu = 2, 3, \cdots, n - 1$, where G_ν is a polynomial. From (9) and (11) it is seen that

$$\begin{aligned}\varphi(z) &= \frac{1}{2}\frac{1 + zg(z)}{1 - zg(z)} \\ &= \frac{1}{2}\frac{1 + a_1 z + g_1 z^2 + \cdots + g_{n-1}z^n + \cdots}{1 - a_1 z - g_1 z^2 - \cdots - g_{n-1}z^n - \cdots} \\ &= \tfrac{1}{2} + a_1 z + a_2 z^2 + \cdots + a_n z^n + \cdots,\end{aligned}$$

where $a_2 = g_1 + a_1^2$, $a_\nu = g_{\nu-1} + H_\nu(a_1, g_1, g_2, \cdots, g_{\nu-2})$ for $\nu = 3, 4, \cdots, n$, H_ν being a polynomial. Substituting the expressions of the g's in terms of the b's, we find

$$\begin{aligned}a_2 &= (1 - a_1\bar{a}_1)b_1 + a_1^2, \\ a_\nu &= (1 - a_1\bar{a}_1)b_{\nu-1} + A_\nu(a_1, \bar{a}_1, b_1, b_2, \cdots, b_{\nu-2})\end{aligned} \tag{21}$$

for $\nu = 3, 4, \cdots, n$, where A_ν is a polynomial. In a similar manner, we obtain from

$$g(z) = \frac{1}{z}\frac{\varphi(z) - \frac{1}{2}}{\varphi(z) + \frac{1}{2}}, \qquad \varphi_1(z) = \frac{1}{2}\frac{1 - a_1 + (1 - \bar{a}_1)g(z)}{1 + a_1 - (1 + \bar{a}_1)g(z)}$$

the formulas

$$\begin{aligned}b_1 &= \frac{1}{1 - a_1\bar{a}_1}(a_2 - a_1^2), \\ b_\nu &= \frac{1}{(1 - a_1\bar{a}_1)^\nu}B_\nu(a_1, \bar{a}_1, a_2, a_3, \cdots, a_{\nu+1})\end{aligned} \tag{22}$$

for $\nu = 2, 3, \cdots, n - 1$, where B_ν is a polynomial.

We may now summarize the preceding results in the statement that the one-to-one correspondence between the points $a_1, a_2, \cdots, a_n$ for which $|a_1| \neq 1$ and the points $a_1, b_1, b_2, \cdots, b_{n-1}$ defined by (21) and (22) is such that when $a_1, a_2, \cdots, a_n$ belongs to K_n, then $b_1, b_2, \cdots, b_{n-1}$ belongs to K_{n-1} and vice versa. Moreover (21) shows that the a's are bounded when this is the case with the b's (in the exceptional case $|a_1| = 1$ it follows from (14) that $|a_2| = 1, \cdots, |a_n| = 1$) so that K_n is bounded when K_{n-1} is bounded, and since this is evidently the case with K_1 defined by (13), we have the result that

The point set K_n is bounded for every n.

From the continuity of the polynomials contained in (21) and (22) it is seen that if $a_{1\mu}, a_{2\mu}, \cdots, a_{n\mu}$ and $a_{1\mu}, b_{1\mu}, \cdots, b_{n-1,\mu}$ correspond for $\mu = 1, 2, \cdots$ $(a_{1\mu} \neq 1)$, and if

$$\lim_{\mu\to\infty} a_{1\mu} = a_1\ (|a_1| \neq 1), \quad \lim a_{2\mu} = a_2, \quad \cdots, \quad \lim a_{n\mu} = a_n,$$
$$\lim b_{1\mu} = b_1, \quad \cdots, \quad \lim b_{n-1,\mu} = b_{n-1},$$

then the points $a_1, a_2, \cdots, a_n$ and $a_1, b_1, b_2, \cdots, b_{n-1}$ also correspond. Assume that a subset of $a_{1\mu}, \cdots, a_{n\mu}$ consists of points belonging to K_n, and another subset of points not belonging to K_n, then the two corresponding subsets of $b_{1\mu}, b_{2\mu}, \cdots, b_{n-1,\mu}$ will and will not belong to K_{n-1} respectively. Hence there corresponds to the boundary point $a_1, a_2, \cdots, a_n$ of K_n (where $|a_1| < 1$) the boundary point $b_1, b_2, \cdots, b_{n-1}$ of K_{n-1} and vice versa. Therefore interior points of K_n and interior points of K_{n-1} also correspond.

We have shown before (see (13)) that all boundary points of K_1 belong to K_1; assuming the same to be true of K_{n-1}, it is also true of K_n. For to a boundary point of K_n for which $|a_1| < 1$ there corresponds a boundary point of K_{n-1} having an associated function $\varphi_1(z)$. From this $\varphi_1(z)$ we form the corresponding $\varphi(z)$ by means of (20), (19), (11) and (9), and this $\varphi(z)$ is associated with the boundary point of K_n from which we started; this boundary point consequently belongs to K_n. Now let $|a_1| = 1$ and make $a_1 = e^{-\alpha i}$, then $\varphi(z)$ is uniquely determined by (14), and it follows that $a_1 = e^{-\alpha i}, a_2 = e^{-2\alpha i}, \cdots, a_n = e^{-n\alpha i}$ is the only point belonging to K_n for which $a_1 = e^{-\alpha i}$. This point $a_1, a_2, \cdots, a_n$ is moreover a boundary point, since in its neighborhood there are points where $|a_1| > 1$ and which therefore do not belong to K_n.

Hence K_n is a perfect point set.

3. **Determination of the boundary of K_n and the corresponding functions $\varphi(z)$.** It will now be shown that any point $a_1, a_2, \cdots, a_n$ on the boundary of K_n determines uniquely the associated $\varphi(z)$ which is of the form

$$\varphi(z) = \tfrac{1}{2}\lambda_1 \frac{e^{\alpha_1 i} + z}{e^{\alpha_1 i} - z} + \tfrac{1}{2}\lambda_2 \frac{e^{\alpha_2 i} + z}{e^{\alpha_2 i} - z} + \cdots + \tfrac{1}{2}\lambda_k \frac{e^{\alpha_k i} + z}{e^{\alpha_k i} - z}, \tag{23}$$

where the α's are all different from each other and

$$0 \leqq \alpha_\nu < 2\pi, \quad \lambda_\nu > 0, \quad \sum \lambda_\nu = 1 \quad (\nu = 1, 2, \cdots, k \text{ and } 1 \leqq k \leqq n) \tag{24}$$

and consequently, expanding both members of (23) in powers of z, that $a_1, a_2, \cdots, a_n$ admit a unique parametric representation

$$a_\nu = \lambda_1 e^{-\nu\alpha_1 i} + \lambda_2 e^{-\nu\alpha_2 i} + \cdots + \lambda_k e^{-\nu\alpha_k i} \quad (\nu = 1, 2, \cdots, n). \tag{25}$$

Conversely, given any α's and λ's satisfying (24), the point $a_1, a_2, \cdots, a_n$ defined by (25) lies on the boundary of K_n (and the associated function is (23)). Since, by the definition of K_n, the point $a_1, a_2, \cdots, a_m$, where $m < n$, belongs to K_m when $a_1, a_2, \cdots, a_n$ belongs to K_n, the significance of the number k is clearly that $a_1, a_2, \cdots, a_m$ is interior to K_m for $m < k$ but on the boundary of K_m for $k \leqq m \leqq n$.

All these statements have been proved for K_1; now we assume them

to hold for K_{n-1} and prove them for K_n as follows. Let $\varphi(z)$ be associated with the point $a_1, a_2, \cdots, a_n$ on the boundary of K_n where $|a_1| < 1$ (when $|a_1| = 1$ our theorem is already proved by (14)); then the $\varphi_1(z)$ derived from $\varphi(z)$ in the manner explained in the preceding paragraph corresponds to a point $b_1, b_2, \cdots, b_{n-1}$ on the boundary of K_{n-1} and by hypothesis we therefore have

$$(26)\qquad \varphi_1(z) = \tfrac{1}{2}\mu_1 \frac{e^{\beta_1 i} + z}{e^{\beta_1 i} - z} + \tfrac{1}{2}\mu_2 \frac{e^{\beta_2 i} + z}{e^{\beta_2 i} - z} + \cdots + \tfrac{1}{2}\mu_{k-1} \frac{e^{\beta_{k-1} i} + z}{e^{\beta_{k-1} i} - z},$$

where all the β's are different and

$$0 \leqq \beta_\nu < 2\pi, \quad \mu_\nu > 0, \quad \sum \mu_\nu = 1,$$
$$(\nu = 1, 2, \cdots, k-1 \text{ and } 1 \leqq k - 1 \leqq n - 1),$$

moreover, this $\varphi_1(z)$ is uniquely determined by $b_1, b_2, \cdots, b_{n-1}$, that is, according to (22), by $a_1, a_2, \cdots, a_n$. From (26), (20), (19), (11) and (9) it follows that $\varphi(z)$ is a rational function of degree not exceeding k, and is uniquely determined by $a_1, a_2, \cdots, a_n$.

Let $\overline{\varphi_1(z)}$ be the conjugate of $\varphi_1(z)$, so that

$$\overline{\varphi_1(z)} = \tfrac{1}{2}\mu_1 \frac{e^{-\beta_1 i} + \bar{z}}{e^{-\beta_1 i} - \bar{z}} + \cdots + \tfrac{1}{2}\mu_{k-1} \frac{e^{-\beta_{k-1} i} + \bar{z}}{e^{-\beta_{k-1} i} - \bar{z}};$$

from

$$\frac{e^{-\beta i} + \bar{z}}{e^{-\beta i} - \bar{z}} = \frac{\dfrac{1}{\bar{z}} + e^{\beta i}}{\dfrac{1}{\bar{z}} - e^{\beta i}}$$

it follows that

$$(27)\qquad \overline{\varphi_1(z)} = -\varphi_1\left(\frac{1}{\bar{z}}\right)$$

and from (20), (19), (11) and (9) successively

$$(28)\qquad \begin{aligned} \overline{f_1(z)} &= \frac{1}{f_1\left(\dfrac{1}{\bar{z}}\right)}, \\ \overline{g(z)} &= \frac{1}{g\left(\dfrac{1}{\bar{z}}\right)}, \\ \overline{f(z)} &= \frac{1}{f\left(\dfrac{1}{\bar{z}}\right)}, \\ \overline{\varphi(z)} &= -\varphi\left(\frac{1}{\bar{z}}\right).* \end{aligned}$$

* The connection of all these equations with Schwarz' principle of reflexion is obvious.

This last equation shows that to a pole z of $\varphi(z)$ inside the unit circle there corresponds a pole $1/\bar{z}$ outside the circle and vice versa, but, by hypothesis, $\varphi(z)$ is holomorphic for $|z| < 1$, and consequently all its poles lie on the unit circle. Let $e^{\alpha i}$ be one of these poles; in its neighborhood we have the expansion

$$\varphi(z) = \tfrac{1}{2}\lambda\left(\frac{e^{\alpha i} + z}{e^{\alpha i} - z}\right)^m + \lambda'\left(\frac{e^{\alpha i} + z}{e^{\alpha i} - z}\right)^{m-1} + \cdots + \lambda^{(m)}\frac{e^{\alpha i} + z}{e^{\alpha i} - z} + P(z - e^{\alpha i})$$

where $\lambda \neq 0$ and P contains positive powers only. Now make

$$z = e^{\alpha i}(1 - \rho e^{\theta i}),$$

then $|z|^2 = 1 - 2\rho\cos\theta + \rho^2$ so that $|z| < 1$ for $-\dfrac{\pi}{2} + \epsilon \leqq \theta \leqq \dfrac{\pi}{2} - \epsilon$ and $0 < \rho < 2\sin\epsilon$, where ϵ is as small as we please. Writing $\lambda = |\lambda| e^{\gamma i}$, $0 \leqq \gamma < 2\pi$, the preceding expansion gives

$$\varphi(z) = |\lambda| e^{(\gamma - m\theta)i} \cdot \frac{2^{m-1}}{\rho^m} + \cdots,$$

and since $\Re\varphi(z) > 0$ for $|z| < 1$, it is necessary that $\cos(\gamma - m\theta) \geqq 0$ for $-\dfrac{\pi}{2} + \epsilon \leqq \theta \leqq \dfrac{\pi}{2} - \epsilon$, that is, letting ϵ approach zero,

$$\cos(\gamma - m\theta) \geqq 0 \qquad \text{for} \qquad -\frac{\pi}{2} \leqq \theta \leqq \frac{\pi}{2}.$$

When θ varies in this interval of length π, and $m \geqq 2$, then $\gamma - m\theta$ varies over more than π, so that $\cos(\gamma - m\theta) < 0$ for some value of θ in the interval. Hence $m = 1$, and $\gamma - \theta$ varies over an interval of length π, which must coincide with the interval where the cosine is positive, that is, the interval from $-\dfrac{\pi}{2}$ to $\dfrac{\pi}{2}$, and consequently $\gamma = 0$. Hence λ is positive, the pole $e^{\alpha i}$ is simple, and since $\varphi(z)$ is of degree $\leqq k$, the number k' of α's cannot exceed k, and consequently $\varphi(z) - \sum_1^{k'} \tfrac{1}{2}\lambda_\nu \dfrac{e^{\alpha_\nu i} + z}{e^{\alpha_\nu i} - z}$ has no poles and therefore equals a constant c. Now, by hypothesis, $\varphi(0) = \tfrac{1}{2}$, whence $\varphi(\infty) = -\tfrac{1}{2}$ by (28); hence $\sum \tfrac{1}{2}\lambda_\nu + c = \tfrac{1}{2}$, $-\sum\tfrac{1}{2}\lambda_\nu + c = -\tfrac{1}{2}$, so that $c = 0$, $\sum\lambda_\nu = 1$, and

$$\varphi(z) = \sum_1^{k'} \tfrac{1}{2}\lambda_\nu \frac{e^{\alpha_\nu i} + z}{e^{\alpha_\nu i} - z}.$$

It remains to show that $k' = k$. From the preceding expression for $\varphi(z)$, we form $g(z)$; using $\sum\lambda_1 = 1$, we find

$$g(z) = \frac{1}{z}\frac{\varphi(z) - \frac{1}{2}}{\varphi(z) + \frac{1}{2}} = \frac{1}{z} \cdot \frac{\sum_1^{k'}\lambda_\nu\left(\dfrac{e^{\alpha_\nu i} + z}{e^{\alpha_\nu i} - z} - 1\right)}{\sum_1^{k'}\lambda_\nu\left(\dfrac{e^{\alpha_\nu i} + z}{e^{\alpha_\nu i} - z} + 1\right)} = \frac{\sum_1^{k'}\dfrac{\lambda_\nu}{e^{\alpha_\nu i} - z}}{\sum_1^{k'}\dfrac{\lambda_1 e^{\alpha_\nu i}}{e^{\alpha_\nu i} - z}},$$

so that the degree of $g(z)$ does not exceed $k' - 1$, and by (19) and (20), the degree $k - 1$ of $\varphi_1(z)$ does not exceed $k' - 1$, or $k \leqq k'$. Since it was shown before that $k' \leqq k$, it follows that $k' = k$, and the first part of our theorem is proved.

To prove the second part, viz., that any α's and λ's satisfying (24) and substituted in (25) yield a point $a_1, a_2, \cdots, a_n$ on the boundary of K_n, we note that since the real part of $\dfrac{e^{\alpha i} + z}{e^{\alpha i} - z}$ is positive for $|z| < 1$, the function (23), formed with any α's and λ's satisfying (24), fulfills all the conditions imposed on $\varphi(z)$. Hence the corresponding $a_1, a_2, \cdots, a_n$, given by (25), belongs to K_n, and all that remains to be shown is that this point lies on the boundary of K_n. We observe first that by (24) and (25) $|a_1| < 1$ unless $k = 1$ and $|a_1| = 1$, which case has been dealt with previously. From (23) and (24), (28) follows, and forming $\varphi_1(z)$ by means of (23), (9), (11), (19) and (20), it is seen that (27) is a consequence of (28). From (27), we conclude that to a pole z of $\varphi_1(z)$ inside the unit circle there corresponds a pole $1/\bar{z}$ outside the circle and vice versa, but $\varphi_1(z)$ being holomorphic for $|z| < 1$, all its poles therefore lie on the unit circle. The real part of $\varphi_1(z)$ being positive for $|z| < 1$, we conclude, by the reasoning previously applied to $\varphi(z)$, that all the poles are simple, that their number does not exceed $k - 1$ (since the degree of $g(z)$ does not exceed $k - 1$), and that we have the expansion

$$\varphi_1(z) = \sum_1^{k'-1} \tfrac{1}{2}\mu_\nu \frac{e^{\beta_\nu i} + z}{e^{\beta_\nu i} - z} + c,$$

where $k' \leqq k$, c is a constant and all $\mu_\nu > 0$. From the way $\varphi_1(z)$ is obtained from $\varphi(z)$ defined by (23), it follows that $\varphi_1(0) = \frac{1}{2}$, $\varphi_1(\infty) = -\frac{1}{2}$, so that $c = 0$, $\sum \mu_\nu = 1$. Hence $\varphi_1(z)$ is of the form (26), and since our theorem is assumed to be proved for K_{n-1}, the point $b_1, b_2, \cdots, b_{n-1}$ to which $\varphi_1(z)$ is associated, lies on the boundary of K_{n-1}. From the correspondence between K_n and K_{n-1}, it now follows that $a_1, a_2, \cdots, a_{n-1}$ lies on the boundary of K_n.

4. **Alternative proof of the results of the preceding paragraph.** The proof now to be presented is as simple as the preceding one and has the advantage of showing in addition that the poles $e^{\alpha_1 i}, \cdots, e^{\alpha_k i}$ of $\varphi(z)$ on one hand, and the poles $e^{\beta_1 i}, \cdots, e^{\beta_{k-1} i}$ of $\varphi_1(z)$ together with $e^{\beta_k i} = \dfrac{1 + \bar{a}_1}{1 + a_1}$ on the other, separate each other on the unit circle (except in a limiting case, where $e^{\alpha_k i}$ and $e^{\beta_k i}$ coincide). Eliminating the intermediate functions from (9), (11), (19) and (20), we find

$$(29)\qquad \varphi(z) = \frac{1}{2}\,\frac{\dfrac{1}{2}\dfrac{1 - \bar{a}_1 - (1 - a_1)z}{1 + \bar{a}_1 - (1 + a_1)z} + \dfrac{1 + \bar{a}_1 + (1 + a_1)z}{1 + \bar{a}_1 - (1 + a_1)z}\varphi_1(z)}{\dfrac{1}{2}\dfrac{1 - \bar{a}_1 + (1 - a_1)z}{1 + \bar{a}_1 - (1 + a_1)z} + \varphi_1(z)}.$$

To prove the first part of our theorem, assume $\varphi(z)$ to be associated with a point $a_1, a_2, \cdots, a_n$ on the boundary of K_n where $|a_1| < 1$; it follows that $\varphi_1(z)$ has the form (26), and consequently that (27) holds. Writing

$$i\psi(z) = \frac{1}{2}\frac{1 - \bar{a}_1 + (1 - a_1)z}{1 + \bar{a}_1 - (1 + a_1)z} + \varphi_1(z),$$

it is seen from (27) that

$$\psi(z) = \psi\left(\frac{1}{\bar{z}}\right),$$

and consequently $\psi(z)$ is real when $|z| = 1$. Making $z = e^{(\beta_\nu + \theta)i}$, where θ is sufficiently small, we now evidently have the following expansions

$$\psi(z) = \mu_\nu \cdot \frac{1}{\theta} + P(\theta)$$

when $e^{\beta_\nu i}$ does not coincide with $e^{\beta_k i} = \dfrac{1 + \bar{a}_1}{1 + a_1}$;

$$\psi(z) = \frac{1 - |a_1|^2}{|1 + a_1|^2}\cdot\frac{1}{\theta} + P(\theta)$$

when $\nu = k$ and $e^{\beta_k i}$ does not coincide with any other $e^{\beta_\nu i}$;

$$\psi(z) = \left(\mu_\nu + \frac{1 - |a_1|^2}{|1 + a_1|^2}\right)\frac{1}{\theta} + P(\theta)$$

when $e^{\beta_\nu i}$ $(\nu < k)$ coincides with $e^{\beta_k i}$. The coefficient of $1/\theta$ being positive in all three cases, it follows that, $e^{\beta_1 i}, \cdots, e^{\beta_{k-1} i}, e^{\beta_k i}$ being arranged in order on the unit circle, the interval between two consecutive ones contains an odd number of zeros of $\psi(z)$. Since the degree of $\psi(z)$ is $k - 1$ or k, according as $e^{\beta_k i}$ does or does not coincide with another $e^{\beta_\nu i}$ where $\nu < k$, it follows that each of these intervals, the number of which is $k - 1$ or k, contains exactly one simple zero of $\psi(z)$, and that this function has no other zeros. By (29), the poles of $\varphi(z)$ are the k zeros of $\psi(z)$ when $e^{\beta_k i}$ does not coincide with any other $e^{\beta_\nu i}$, but when this is the case, then $e^{\beta_k i}$ is a simple pole of $\varphi(z)$, the other poles being the $k - 1$ zeros of $\psi(z)$. Consequently

$$\varphi(z) = \sum_{\nu=1}^{k} \tfrac{1}{2}\lambda_\nu \frac{e^{\alpha_\nu i} + z}{e^{\alpha_\nu i} - z} + c$$

where $e^{\alpha_1 i}, \cdots, e^{\alpha_k i}$ separate and are separated by $e^{\beta_1 i}, \cdots, e^{\beta_{k-1} i}, e^{\beta_k i}$, except when $e^{\beta_k i}$ coincides with another $e^{\beta_\nu i}$, in which case one $e^{\alpha_\nu i}$, say

$e^{\alpha_k i}$, coincides with $e^{\beta_k i}$ and the $k-1$ other $e^{\alpha_\nu i}$ separate and are separated by $e^{\beta_1 i}, \cdots, e^{\beta_{k-1} i}$. The proof that $\sum_1^k \lambda_\nu = 1$ and $c = 0$ is the same as in the preceding paragraph, and we may either use the method given there to show that all λ_ν are positive, or we may use $\varphi(z) = -\varphi\left(\frac{1}{z}\right)$ to show that all λ's are real, and then make $z = \rho e^{\alpha_\nu i}$ and let ρ approach unity to conclude that $\lambda_\nu > 0$.

To prove the second part of our theorem, we assume $\varphi(z)$ to be of the form (23) with given α's and λ's satisfying (24). Then evidently (28) is true. Compute the corresponding $\varphi_1(z)$; we find from (29)

$$(30) \qquad \varphi_1(z) = \frac{1}{2}\,\frac{\dfrac{1}{2}\,\dfrac{1-\bar{a}_1-(1-a_1)z}{1+\bar{a}_1-(1+a_1)z} - \dfrac{1-\bar{a}_1+(1-a_1)z}{1+\bar{a}_1-(1+a_1)z}\,\varphi(z)}{\varphi(z) - \dfrac{1}{2}\,\dfrac{1+\bar{a}_1+(1+a_1)z}{1+\bar{a}_1-(1+a_1)z}}.$$

Writing

$$i\psi_1(z) = \varphi(z) - \frac{1}{2}\,\frac{1+\bar{a}_1+(1+a_1)z}{1+\bar{a}_1-(1+a_1)z},$$

it follows from (28) that $\psi_1(z)$ is real for $|z| = 1$. The degree of $\psi_1(z)$ is k or $k+1$ according as $e^{\beta_k i} = \dfrac{1+\bar{a}_1}{1+a_1}$ does or does not coincide with any of the $e^{\alpha_\nu i}$, and it is evident at once that $\psi_1(z) = 0$ for $z = 0$ and $z = \infty$, so that there remain $k-2$ or $k-1$ zeros respectively to be located. Making $z = e^{(\alpha_\nu+\theta)i}$, we have the expansions

$$\psi_1(z) = \lambda_\nu \cdot \frac{1}{\theta} + P(\theta)$$

when $e^{\alpha_\nu i}$ does not coincide with $e^{\beta_k i}$,

$$\psi_1(z) = -(1-\lambda_\nu)\cdot\frac{1}{\theta} + P(\theta)$$

when $e^{\alpha_\nu i}$ coincides with $e^{\beta_k i}$, and for $z = e^{(\beta_k+\theta)i}$

$$\psi_1(z) = -\frac{1}{\theta} + P(\theta)$$

when $e^{\beta_k i}$ does not coincide with any $e^{\alpha_\nu i}$. Hence arranging $e^{\alpha_1 i}, \cdots, e^{\alpha_k i}$ and $e^{\beta_k i}$ in order on the unit circle, obtaining k or $k+1$ intervals according as there is coincidence or not, it follows that the two intervals adjacent to $e^{\beta_k i}$ contain an even number of zeros of $\psi(z)$, the remaining $k-2$ or $k-1$ intervals an odd number. But there were exactly $k-2$ or $k-1$ zeros to be located, and it follows that they are all simple and situated one in

each of the intervals not adjacent to $e^{\beta_\nu i}$. Now reasoning on the poles of (30) as before on those of (29), we find that $\varphi_1(z)$ has the form

$$\varphi_1(z) = \sum_{\nu=1}^{k-1} \tfrac{1}{2}\mu_\nu \frac{e^{\beta_\nu i} + z}{e^{\beta_\nu i} - z} + c$$

with the separation of the $e^{\alpha_\nu i}$ and the $e^{\beta_\nu i}$ found previously; from (30) it is seen at once that $\varphi_1(0) = \frac{1}{2}$, $\varphi_1(\infty) = -\frac{1}{2}$, hence $\sum_1^{k-1} \mu_\nu = 1$, $c = 0$. We prove as for $\varphi(z)$ that all μ's are positive, and hence $\varphi_1(z)$ has the form (26), being therefore associated with a point $b_1, \cdots, b_{n-1}$ on the boundary of K_{n-1}, and it finally follows that $a_1, a_2, \cdots, a_n$ is on the boundary of K_n.

5. **Proof that K_n is a convex solid, and parametric representation of its interior points.** Let $\varphi_1(z)$ and $\varphi_2(z)$ be two functions associated with the points $a_1', a_2', \cdots, a_n'$ and $a_1'', a_2'', \cdots, a_n''$ both belonging to K_n. The function $\varphi(z) = (1 - t)\varphi_1(z) + t\varphi_2(z)$ where $0 \leqq t \leqq 1$ evidently is holomorphic and of positive real part for $|z| < 1$, and $\varphi(0) = \frac{1}{2}$. Consequently, $\varphi(z)$ is associated with the point $a_1, a_2, \cdots, a_n$, where $a_\nu = (1 - t)a_\nu' + ta_\nu''$, so that this point also belongs to K_n. Therefore K_n is a convex point set, and being perfect, bounded, and containing a $2n$-dimensional neighborhood of the origin as interior points, K_n is a convex $2n$-dimensional solid according to Minkowski's definition.

It is readily seen that when $a_1, a_2, \cdots, a_n$ belongs to K_n, then $ta_1, ta_2, \cdots, ta_n$ is an interior point for $0 \leqq t < 1$. In fact, there exists a neighborhood ϵ of the origin such that all its points belong to K_n; to any point $a_1', a_2', \cdots, a_n'$ such that $|a_\nu' - ta_\nu| < \epsilon(1 - t)$ for $\nu = 1, 2, \cdots, n$ we adjoin another $a_1'', a_2'', \cdots, a_n''$ by the equations $a_\nu' = (1 - t)a_\nu'' + ta_\nu$. It follows that $(1 - t)|a_\nu''| = |a_\nu' - ta_\nu| < \epsilon(1 - t)$ or $|a_\nu''| < \epsilon$, so that $a_1'', a_2'', \cdots, a_n''$ belongs to K_n, and consequently $a_1', a_2', \cdots, a_n'$ also belongs to K_n, since it lies on the segment joining $a_1'', a_2'', \cdots, a_n''$ and $a_1, a_2, \cdots, a_n$. Thus the neighborhood $(1 - t)\epsilon$ of $ta_1, ta_2, \cdots, ta_n$ belongs to K_n, and $ta_1, ta_2, \cdots, ta_n$ is therefore an interior point.

This result may also be expressed as follows: when $a_1, a_2, \cdots, a_n$ is a point interior to K_n but distinct from the origin, then there exists one and only one t, where $0 < t < 1$, such that the point $\frac{a_1}{t}, \frac{a_2}{t}, \cdots, \frac{a_n}{t}$ is on the boundary of K_n. By (25) and (24), this boundary point has the unique parametric representation

$$\frac{a_\nu}{t} = \lambda_1' e^{-\nu\alpha_1 i} + \lambda_2' e^{-\nu\alpha_2 i} + \cdots + \lambda_k' e^{-\nu\alpha_k i} \quad (\nu = 1, 2, \cdots, n)$$

with $0 \leqq \alpha_\nu < 2\pi$, all α's different, $\lambda_\nu' > 0$, $\sum \lambda_\nu' = 1$ $(\nu = 1, 2, \cdots, k)$ and $1 \leqq k \leqq n$. Writing $t\lambda_\nu' = \lambda_\nu$, it is seen that the interior point

$a_1, a_2, \cdots, a_n$ has the unique parametric representation

$$(25') \qquad a_\nu = \lambda_1 e^{-\nu\alpha_1 i} + \lambda_2 e^{-\nu\alpha_2 i} + \cdots + \lambda_k e^{-\nu\alpha_k i} \quad (\nu = 1, 2, \cdots, n)$$

with the α's all different and

$$(24') \qquad a \leqq \alpha_\nu < 2\pi, \quad \lambda_\nu > 0, \quad \sum \lambda_\nu < 1 \qquad (\nu = 1, 2, \cdots, k \text{ and } 1 \leqq k \leqq n).$$

Making all λ's equal to zero, this result holds also for the origin. To prove that conversely the point defined by (25′), from given α's and λ's satisfying (24′), is interior to K_n, we remark that

$$(23') \quad \varphi(z) = \tfrac{1}{2}\lambda_0 + \tfrac{1}{2}\lambda_1 \frac{e^{\alpha_1 i} + z}{e^{\alpha_1 i} - z} + \tfrac{1}{2}\lambda_2 \frac{e^{\alpha_2 i} + z}{e^{\alpha_2 i} - z} + \cdots + \tfrac{1}{2}\lambda_k \frac{e^{\alpha_k i} + z}{e^{\alpha_k i} - z},$$

where $\lambda_0 = 1 - \lambda_1 - \lambda_2 - \cdots - \lambda_k > 0$, is evidently a φ-function associated with the point $a_1, a_2, \cdots, a_n$ defined by (25′) which therefore belongs to K_n. The point is an interior one, since if it were on the boundary, (23) would give $\varphi(z)$ uniquely and in the form

$$\varphi(z) = \tfrac{1}{2}\lambda_1' \frac{e^{\alpha_1' i} + z}{e^{\alpha_1' i} - z} + \cdots + \tfrac{1}{2}\lambda_m' \frac{e^{\alpha_m' i} + z}{e^{\alpha_m' i} - z},$$

which cannot coincide with (23′) unless $m = k$, $\alpha_\nu' = \alpha_\nu$, $\lambda_\nu' = \lambda_\nu$ and consequently $\lambda_0 = 0$ contrary to (24′). It should be noted that (23′) is not the only $\varphi(z)$ associated with the interior point $a_1, a_2, \cdots, a_n$.

6. **The characterization of K_n by algebraic inequalities involving $a_1, a_2, \cdots, a_n$ and their conjugates.** These inequalities are already stated in (5) and (6), the D's being defined by (4) and $D_0 = 1$. For $n = 1$, these inequalities reduce to $D_1 > 0$ in the interior and $D_1 = 0$ on the boundary of K_1, and since (4) gives $D_1(a_1) = 1 - a_1\bar{a}_1 = 1 - |a_1|^2$, the desired result is obtained immediately by comparison with (13). From what has been said before regarding the correspondence between K_n and K_{n-1}, it is obvious that the inequalities (5) and (6) follow in the general case by complete induction from the identity

$$(31) \qquad D_m(a_1, a_2, \cdots, a_m) = (1 - a_1\bar{a}_1)^m D_{m-1}(b_1, b_2, \cdots, b_{m-1}), \qquad m = 1, 2, \cdots, n,$$

which we shall now proceed to prove. We begin by showing that when $a_1, a_2, \cdots, a_n$ is on the boundary of K_n, then $D_m(a_1, a_2, \cdots, a_m) = 0$ for $m \geqq k$, where k is the integer occurring in the parametric representation (25). In fact, by (25) and (24), the element in the pth column and qth row of the determinant (4) is seen at once to be

$$\sum_{\nu=1}^{k} \lambda_\nu e^{(q-p)\alpha_\nu i},$$

and expanding the determinant in powers of the λ's, we find

$$D_m(a_1, a_2, \cdots, a_m) = \sum_{\nu_1, \nu_2, \cdots, \nu_{m+1} = 1, 2, \cdots, k} \lambda_{\nu_1}\lambda_{\nu_2}\cdots\lambda_{\nu_{m+1}} \left| e^{(q-p)\alpha_{\nu_p} i} \right|_{p, q = 1, 2, \cdots, m+1}.$$

Since there are only $k < m + 1$ different α's, any one of the determinants to the right contains the same α_ν in two columns, say the pth and rth, and the latter column is obtained from the former by multiplication by $e^{(p-r)\alpha_\nu}p^t$, so that the determinant vanishes, and consequently

$$D_m(a_1, a_2, \cdots, a_m) = 0 \qquad \text{for} \qquad k \leqq m \leqq n \tag{32}$$

when $a_1, a_2, \cdots, a_n$ is on the boundary of K_n.

Next, we observe that (31) is obtained at once by direct calculation of the determinant (4) for $m = 1$ and $m = 2$, using the expression (21) for a_2 in the latter case.

Now assume the inequalities (5) and (6) proved for $K_1, K_2, \cdots, K_{n-1}$, and that the identity (31) holds for $m = 1, 2, \cdots, n - 1$. Assume $|a_1| < 1$ and that $b_1, b_2, \cdots, b_{n-1}$ satisfies $D_{n-1}(b_1, b_2, \cdots, b_{n-1}) = 0$ and the further conditions

$$D_1(b_1) > 0, \quad D_2(b_1, b_2) > 0, \quad \cdots, \quad D_{n-2}(b_1, b_2, \cdots, b_{n-2}) > 0. \tag{33}$$

Then $b_1, b_2, \cdots, b_{n-1}$ is on the boundary of K_{n-1} (but $b_1, b_2, \cdots, b_{n-2}$ interior to K_{n-2}) and consequently, calculating $a_2, \cdots, a_n$ from (21), the point $a_1, a_2, \cdots, a_n$ is on the boundary of K_n, so that $D_n(a_1, a_2, \cdots, a_n) = 0$ by (32). In other words, taking arbitrary fixed values of $b_1, b_2, \cdots, b_{n-2}$ satisfying (33) and a variable b_{n-1}, and calculating $a_1, a_2, \cdots, a_n$ by (21), then $D_n(a_1, a_2, \cdots, a_n)$ becomes a polynomial in the two variables b_{n-1} and $\bar{b}_{n-1}$, which vanishes whenever the polynomial in the same two variables $D_{n-1}(b_1, b_2, \cdots, b_{n-1})$ vanishes. Consequently the former polynomial is divisible by the latter:

$$\begin{aligned} D_n(a_1, a_2, \cdots, a_n) & \\ &= \psi(a_1, \bar{a}_1, b_1, \bar{b}_1, \cdots, b_{n-1}, \bar{b}_{n-1}) D_{n-1}(b_1, b_2, \cdots, b_{n-1}) \end{aligned} \tag{34}$$

where ψ is a polynomial in b_{n-1} and $\bar{b}_{n-1}$. By (4), D_n is linear in each of the two variables a_n and $\bar{a}_n$, the coefficient of $a_n\bar{a}_n$ being $- D_{n-2}(a_1, a_2, \cdots, a_{n-2})$, hence using (21), we see that D_n is linear in each of the variables b_{n-1} and $\bar{b}_{n-1}$, the coefficient of $b_{n-1}\bar{b}_{n-1}$ being $- (1 - a_1\bar{a}_1)^2 D_{n-2}(a_1, a_2, \cdots, a_{n-2})$. The coefficient of $b_{n-1}\bar{b}_{n-1}$ in $D_{n-1}(b_1, b_2, \cdots, b_{n-1})$ being $- D_{n-3}(b_1, b_2, \cdots, b_{n-3})$, it follows from (34) that ψ cannot contain b_{n-1} or $\bar{b}_{n-1}$, and comparing coefficients of $b_{n-1}\bar{b}_{n-1}$ on both sides, it is seen that

$$(1 - a_1\bar{a}_1)^2 D_{n-2}(a_1, a_2, \cdots, a_{n-2}) = \psi \cdot D_{n-3}(b_1, b_2, \cdots, b_{n-3}).$$

By hypothesis, (31) is proved for $m = n - 2$, so that

$$D_{n-2}(a_1, a_2, \cdots, a_{n-2}) = (1 - a_1\bar{a}_1)^{n-2} D_{n-3}(b_1, b_2, \cdots, b_{n-3});$$

introducing this in the preceding equation and dividing by D_{n-3} which does not vanish by (33), we find $\psi = (1 - a_1\bar{a}_1)^n$ and (31) is proved for $m = n$ (being an algebraic identity, it evidently also holds when the conditions (33) are not satisfied). The induction proof of the inequalities (5) and (6) is now complete.

New York City,
August 10, 1921.

ALGEBRAIC SURFACES, THEIR CYCLES AND INTEGRALS. A CORRECTION.

By S. Lefschetz.

1. In a paper under the same title (these Annals, vol. 21, 1920), whose notations shall be used here, I gave a treatment of the topology of algebraic surfaces. My first object here is to call attention to two incorrect proofs kindly pointed out to me by J. W. Alexander. I then propose to give analytical proofs in place of one of these.

The defective proofs refer to two theorems on linear cycles, correct themselves, the second part of the theorem in No. 8 and the theorem in No. 9.—That the proofs could not hold was discovered by Alexander by means of the "quasi-algebraic" manifold

$$z^2 = x(x-a)(x-b)(x-y); \qquad |y|,\ |a|,\ |b| < 1;$$
$$z^2 = x(x-a)(x-b)\left(x - \frac{1}{\bar{y}}\right); \qquad |y| \geqq 1, \quad (\bar{y} \text{ conjugate of } y).$$

This manifold behaves in many respects like an algebraic surface. However its linear index $R_1 = 1$, whereas by the reasoning of the paper (No. 10), apparently applicable here, it should be even.—Modifications necessitated in the discussion have fortunately resulted in the discovery of new and very interesting properties. The whole question will be treated elsewhere at length in the near future.* Suffice to say for the present that the solution of the difficulties was found: (*a*) For No. 8 in a new proof involving the fact that the curve H_y belongs to a linear system ∞^2 at least. (*b*) For No. 9 in a further study of the linear cycles of the curve based on the following added precision to the Picard theorem given in No. 11 regarding the behavior of a cycle Γ_1 of H_y when y is near a critical point a_i: *The increment of* Γ_1 *when* y *turns around* a_i *is equal to* $(\Gamma_1\delta_i)\cdot\delta_i$. This seemingly unimportant point proved of the utmost value.

2. Of the theorem in No. 8 there is a very simple analytical proof. The question is to show that *if* Γ_1 *is invariant and bounds on the surface so does its locus* Γ_3 *when* y *varies*. It suffices to show that Γ_1 itself bounds on H_y. Now it has been proved independently of our theorem (loc. cit., No. 15) that Γ_1, zero cycle of the surface, is related by a homology to the vanishing cycles δ_i. Hence all reduces to showing that an invariant sum of (δ)'s bounds on H_y. Let $\Gamma_1 = \sum \lambda_i\delta_i$ be such an invariant cycle. There is an integral of total differentials of the second kind with a period $+1$ relatively to Γ_1 (Picard). But its periods relatively to the (δ)'s are all zero since these cycles are deformable into points of the surface. Hence the period relatively to Γ_1 must also vanish, a contradiction which proves the theorem.

* In a monograph to appear in the Borel Series.

V. 22 c 23

SEPTEMBER 1920

math.

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

ORMOND STONE
L. P. EISENHART
OSWALD VEBLEN
PIERRE BOUTROUX
T. H. GRONWALL
J. H. M. WEDDERBURN

WITH THE COÖPERATION OF

J. W. ALEXANDER
H. BLUMBERG
A. A. BENNETT
G. A. PFEIFFER
J. K. WHITTEMORE

PUBLISHED UNDER THE AUSPICES OF PRINCETON UNIVERSITY

SECOND SERIES, VOL. 22, NO. 1

LANCASTER, PA., AND PRINCETON, N. J.

An agreement has been made between the Mathematical Association of America and the Annals of Mathematics in accordance with which the Association contributes to the support of the Annals; the Annals has enlarged its volume by 100 pages which are devoted to expository and historical articles in so far as suitable material of this sort is available; and the subscription price to individual members of the Association is one-half of the regular price. The regular subscription price of the Annals beginning with Volume 19 is $3.00 a volume.

Copies of the following memoirs can be obtained by addressing THE ANNALS OF MATHEMATICS, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. JENSEN. Authorized translation with additional notes by T. H. GRONWALL. 43 pages. Price 50 cents.

The gamma function in the integral calculus. By T. H. GRONWALL 89 pages. 90 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. DICKSON. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. OSGOOD. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. JENSEN. Authorized translation from the Danish by T. H. GRONWALL. 29 pages. Price 40 cents.

CONTENTS

PAGE

On multiform functions defined by differential equations of the first order. By P. Boutroux 1

Hermitian Metrics. By J. L. Coolidge 11

On the expansion of certain analytic functions in series. By R. D. Carmichael 29

Notes on the cyclic quadrilateral. By F. V. Morley. 35

Note on the preceding paper. By F. Morley 43

Qualitative properties of the ballistic trajectory. By T. H. Gronwall . . . 44

ANNALS OF MATHEMATICS

Published in September, December, March and June, at Lancaster, Pa. and Princeton, N. J., U. S. A. Subscription price, $3 a volume (four numbers) in advance. Single numbers, $1.00. All drafts and money orders should be made payable to The Annals of Mathematics, Princeton, New Jersey.

Manuscripts and other communications should be addressed to The Annals of Mathematics, Princeton, N. J., U. S. A.

Authors receive gratis 25 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

Press of The New Era Printing Company, Lancaster, Pa.

Math

DECEMBER, 1920

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

ORMOND STONE — PIERRE BOUTROUX
L. P. EISENHART — T. H. GRONWALL
OSWALD VEBLEN — J. H. M. WEDDERBURN

WITH THE COÖPERATION OF

J. W. ALEXANDER — A. A. BENNETT
H. BLUMBERG — G. A. PFEIFFER
J. K. WHITTEMORE

PUBLISHED UNDER THE AUSPICES OF PRINCETON UNIVERSITY

SECOND SERIES, VOL. 22, NO. 2

LANCASTER, PA., AND PRINCETON, N. J.

An agreement has been made between the Mathematical Association of America and the Annals of Mathematics in accordance with which the Association contributes to the support of the Annals; the Annals has enlarged its volume by 100 pages which are devoted to expository and historical articles in so far as suitable material of this sort is available; and the subscription price to individual members of the Association is one-half of the regular price. The regular subscription price of the Annals beginning with Volume 19 is $3.00 a volume.

Copies of the following memoirs can be obtained by addressing THE ANNALS OF MATHEMATICS, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. JENSEN. Authorized translation with additional notes by T. H. GRONWALL. 43 pages. Price 50 cents.

The gamma function in the integral calculus. By T. H. GRONWALL 89 pages. 90 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. DICKSON. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. OSGOOD. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. JENSEN. Authorized translation from the Danish by T. H. GRONWALL. 29 pages. Price 40 cents.

CONTENTS

PAGE

The mean of a functional of arbitrary elements. By N. Wiener 66

On certain determinants associated with transformations employed in thermodynamics. By J. F. Trevor 73

The permanent gravitational field in the Einstein theory. By L. P. Eisenhart . 86

On the structure of finite continuous groups with a finite number of exceptional infinitesimal transformations. By S. D. Zeldin 95

Conformal mapping of a family of real conics upon another. By T. H. Gronwall. 101

On the location of the roots of the derivative of a polynomial. By J. L. Walsh. 128

ANNALS OF MATHEMATICS

Published in September, December, March and June, at Lancaster Pa. and Princeton, N. J., U. S. A. Subscription price, $3 a volume (four numbers) in advance. Single numbers, $1.00. All drafts and money orders should be made payable to The Annals of Mathematics, Princeton, New Jersey.

Manuscripts and other communications should be addressed to The Annals of Mathematics, Princeton, N. J., U. S. A.

Authors receive gratis 25 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

Press of The New Era Printing Company, Lancaster, Pa.

Stanford Library
FEB 28 1921
Math

MARCH. 1921

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

ORMOND STONE — PIERRE BOUTROUX
L. P. EISENHART — T. H. GRONWALL
OSWALD VEBLEN — J. H. M. WEDDERBURN

WITH THE COÖPERATION OF

J. W. ALEXANDER — A. A. BENNETT
H. BLUMBERG — G. A. PFEIFFER
J. K. WHITTEMORE

PUBLISHED UNDER THE AUSPICES OF PRINCETON UNIVERSITY

SECOND SERIES, VOL. 22, NO. 3

LANCASTER, PA., AND PRINCETON, N. J.

An agreement has been made between the Mathematical Association of America and the Annals of Mathematics in accordance with which the Association contributes to the support of the Annals; the Annals has enlarged its volume by 100 pages which are devoted to expository and historical articles in so far as suitable material of this sort is available; and the subscription price to individual members of the Association is one-half of the regular price. The regular subscription price of the Annals beginning with Volume 19 is $3.00 a volume.

Copies of the following memoirs can be obtained by addressing THE ANNALS OF MATHEMATICS, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. JENSEN. Authorized translation with additional notes by T. H. GRONWALL. 43 pages. Price 50 cents.

The gamma function in the integral calculus. By T. H. GRONWALL 89 pages. 90 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. DICKSON. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. OSGOOD. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. JENSEN. Authorized translation from the Danish by T. H. GRONWALL. 29 pages. Price 40 cents.

CONTENTS

PAGE

The asymptotic expansion of the Sturm-Liouville functions. By F. H. MURRAY 145

On the conformal mapping of a region into a part of itself. By J. F. RITT . . 157

Conjugate nets R and their conformations. By L. P. EISENHART 161

The application of modern theories of integration to the solution of differential equations. By T. C. FRY 182

ANNALS OF MATHEMATICS

Published in September, December, March and June, at Lancaster Pa. and Princeton, N. J., U. S. A. Subscription price, $3 a volume (four numbers) in advance. Single numbers, $1.00. All drafts and money orders should be made payable to THE ANNALS OF MATHEMATICS, Princeton, New Jersey.

Manuscripts and other communications should be addressed to THE ANNALS OF MATHEMATICS, Princeton, N. J., U. S. A.

Authors receive gratis 25 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

PRESS OF THE NEW ERA PRINTING COMPANY, LANCASTER, PA.

JUNE, 1921

Stanford Library
SEP 12 1921
Math.

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

ORMOND STONE
PIERRE BOUTROUX
L. P. EISENHART
T. H. GRONWALL
OSWALD VEBLEN
J. H. M. WEDDERBURN

WITH THE COÖPERATION OF

J. W. ALEXANDER
A. A. BENNETT
H. BLUMBERG
G. A. PFEIFFER
J. K. WHITTEMORE

PUBLISHED UNDER THE AUSPICES OF PRINCETON UNIVERSITY

SECOND SERIES, VOL. 22, NO. 4

LANCASTER, PA., AND PRINCETON, N. J.

According to an agreement between the Mathematical Association of America and the editors of the Annals of Mathematics, the Association contributes to the support of the Annals, and the Annals is supplied to individual members of the Association at one half of the regular price. In consequence of this agreement the volume of the Annals was increased by 100 pages which are devoted to expository and historical articles in so far as suitable articles of this class are obtainable. Thus far the editors have not received enough such articles to fill the space available, and therefore wish to call the attention of authors to this lack and to the fact that as long as the shortage continues expository or historical articles of sufficient merit will receive prompt publication.

A number of the expository articles which have already been published are available in separate form and are listed for sale on the inside of the back cover of this number of the Annals. The regular subscription price of the Annals is $3.00 a volume.

Copies of the following memoirs can be obtained by addressing The Annals of Mathematics, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. Jensen. Authorized translation with additional notes by T. H. Gronwall. 43 pages. Price 50 cents.

The gamma function in the integral calculus. By T. H. Gronwall 89 pages. 90 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. Dickson. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. Osgood. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. Jensen. Authorized translation from the Danish by T. H. Gronwall. 29 pages. Price 40 cents.

CONTENTS

PAGE

An Analytical Solution of Biot's Problem. . . . By Tsuruichi Hayashi 213

Minimal Surfaces Containing Straight Lines. . By James K. Whittemore 217

An Extension of Green's Lemma to the Case of a Rectifiable Boundary. By Edward B. Van Vleck 226

Periodic Conjugate Nets. . . . . By Edward S. Hammond 238

On the Transformation of Convex Point Sets. . . . . . By J. L. Walsh 262

ANNALS OF MATHEMATICS

Published in September, December, March and June, at Lancaster Pa. and Princeton, N. J., U. S. A. Subscription price, $3 a volume (four numbers) in advance. Single numbers, $1.00. All drafts and money orders should be made payable to The Annals of Mathematics, Princeton, New Jersey.

Manuscripts and other communications should be addressed to The Annals of Mathematics, Princeton, N. J., U. S. A.

Authors receive gratis 25 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

Press of The New Era Printing Company, Lancaster, Pa.

Stanford Library
AUG 7 1922
math.

SEPTEMBER, 1921

ANNALS OF MATHEMATICS

(Founded by Ormond Stone)

EDITED BY

ORMOND STONE | PIERRE BOUTROUX
L. P. EISENHART | T. H. GRONWALL
OSWALD VEBLEN | J. H. M. WEDDERBURN

WITH THE COÖPERATION OF

J. W. ALEXANDER | A. A. BENNETT
H. BLUMBERG | G. A. PFEIFFER
J. K. WHITTEMORE

PUBLISHED BY THE

PRINCETON UNIVERSITY PRESS

Second Series, Vol. 23, No. 1

LANCASTER, PA., AND PRINCETON, N. J.

1922

According to an agreement between the Mathematical Association of America and the editors of the Annals of Mathematics, the Association contributes to the support of the Annals, and the Annals is supplied to individual members of the Association at one half of the regular price. In consequence of this agreement the volume of the Annals was increased by 100 pages, which are devoted to expository and historical articles in so far as suitable articles of this class are obtainable. Thus far the editors have not received enough such articles to fill the space available, and therefore wish to call the attention of authors to this lack and to the fact that, as long as the shortage continues, expository or historical articles of sufficient merit will receive prompt publication.

A number of the expository articles which have already been published are available in separate form and are listed for sale on the inside of the back cover of this number of the Annals. The regular subscription price of the Annals is $3.00 a volume.

Copies of the following memoirs can be obtained by addressing The Annals of Mathematics, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. Jensen. Authorized translation with additional notes by T. H. Gronwall. 43 pages. Price 50 cents.

The gamma function in the integral calculus. By T. H. Gronwall. 89 pages. 90 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. Dickson. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. Osgood. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. Jensen. Authorized translation from the Danish by T. H. Gronwall. 29 pages. Price 40 cents.

CONTENTS

Page

On matrices whose elements are integers. By O. VEBLEN and P. FRANKLIN . 1

An algorism for differential invariant theory. By O. E. GLENN 16

The general theory of cyclic-harmonic curves. By R. E. MORITZ 29

More theorems on the complete quadrilateral. By J. W. CLAWSON 40

A theorem on cross-ratios in the geometry of inversion. By J. L. WALSH . . 45

The condition for an isothermal family on a surface. By J. K. WHITTEMORE 52

The reversion of class number relations and the total representation of integers as sums of squares or triangular numbers. By E. T. BELL . . 56

Note on the term maximal subgroup. By G. A. MILLER 68

Reducible cubic forms expressible rationally as determinants. By L. E. DICKSON 70

Note on the Picard method of successive approximations. By D. JACKSON . 75

A fundamental system of covariants of the ternary cubic form. By L. E. DICKSON 78

The modular theory of polyadic numbers. By A. A. BENNETT. 83

Some analogies in matric theory. By A. A. BENNETT 91

Generalized conjugate matrices. By P. Franklin. 97

ANNALS OF MATHEMATICS

Published in September, December, March and June at Lancaster, Pa., and Princeton, N. J. Subscription price, $3 a volume (four numbers) in advance. Single copies $1.00. Subscriptions, orders for back numbers, and changes of address should be sent to the PRINCETON UNIVERSITY PRESS, PRINCETON, NEW JERSEY.

Manuscripts and all editorial correspondence should be addressed to THE ANNALS OF MATHEMATICS, P. O. BOX 53, PRINCETON, NEW JERSEY. Manuscripts should be typewritten, with the exception of formulæ, and must be in final form with all references filled in.

Authors receive gratis 25 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

PRESS OF
THE NEW ERA PRINTING COMPANY
LANCASTER, PA.

Stanford Library
NOV 20 1922
Math

DECEMBER, 1921

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

ORMOND STONE — J. W. ALEXANDER
L. P. EISENHART — T. H. GRONWALL
OSWALD VEBLEN — J. H. M. WEDDERBURN

WITH THE COÖPERATION OF

A. A. BENNETT — H. BLUMBERG
G. A. PFEIFFER — J. K. WHITTEMORE

PUBLISHED BY THE

PRINCETON UNIVERSITY PRESS

SECOND SERIES, VOL. 23, NO. 2

LANCASTER, PA., AND PRINCETON, N. J.

1922

PIERRE BOUTROUX

JUNE 25, 1922

According to an agreement between the Mathematical Association of America and the editors of the Annals of Mathematics, the Association contributes to the support of the Annals, and the Annals is supplied to individual members of the Association at one half of the regular price. In consequence of this agreement the volume of the Annals was increased by 100 pages, which are devoted to expository and historical articles in so far as suitable articles of this class are obtainable. Thus far the editors have not received enough such articles to fill the space available, and therefore wish to call the attention of authors to this lack and to the fact that, as long as the shortage continues, expository or historical articles of sufficient merit will receive prompt publication.

A number of the expository articles which have already been published are available in separate form and are listed for sale on the inside of the back cover of this number of the Annals. The regular subscription price of the Annals is $3.00 a volume.

Entered at the postoffice at Lancaster, Pa., as second-class matter under the Act of March 3, 1879.

Copies of the following memoirs can be obtained by addressing THE ANNALS OF MATHEMATICS, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. JENSEN. Authorized translation with additional notes by T. H. GRONWALL. 43 pages. Price 50 cents.

The gamma function in the integral calculus. By T. H. GRONWALL. 89 pages. 90 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. DICKSON. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. OSGOOD. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. JENSEN. Authorized translation from the Danish by T. H. GRONWALL. 29 pages. Price 40 cents.

CONTENTS

Page

Transformations of trajectories on a surface. By J. LIPKA 101

On the structure of finite continuous groups with one two-parameter invariant subgroup. By S. D. ZELDIN 112

On the simplification of the structure of finite continuous groups with more than one two-parameter invariant subgroup. By S. D. ZELDIN . . . 118

The automorphic transformation of a bilinear form. By J. H. M. WEDDERBURN 122

A direct determination of the minimum area between a curve and its caustic. By O. DUNKEL 135

The Poisson integral and an analytic function on its circle of convergence. By A. ARWIN 141

Systems of circuits on two-dimensional manifolds. By H. R. BRAHANA . . 144

Two generalizations of the Stieltjes integral. By P. J. DANIELL 169

ANNALS OF MATHEMATICS

Published in September, December, March and June at Lancaster, Pa., and Princeton, N. J. Subscription price, $3 a volume (four numbers) in advance. Single copies $1.00. Subscriptions, orders for back numbers, and changes of address should be sent to the PRINCETON UNIVERSITY PRESS, PRINCETON, NEW JERSEY.

Manuscripts and all editorial correspondence should be addressed to THE ANNALS OF MATHEMATICS, P. O. BOX 53, PRINCETON, NEW JERSEY. Manuscripts should be typewritten, with the exception of formulæ, and must be in final form with all references filled in.

Authors receive gratis 25 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

PRESS OF
THE NEW ERA PRINTING COMPANY
LANCASTER, PA.

Stanford Library
MAY 2 1923
Math

MARCH, 1922

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

ORMOND STONE | J. W. ALEXANDER
L. P. EISENHART | T. H. GRONWALL
OSWALD VEBLEN | J. H. M. WEDDERBURN

WITH THE COÖPERATION OF

A. A. BENNETT | H. BLUMBERG
G. A. PFEIFFER | J. K. WHITTEMORE

PUBLISHED BY THE

PRINCETON UNIVERSITY PRESS

SECOND SERIES, VOL. 23, NO. 3

LANCASTER, PA., AND PRINCETON, N. J.

1923

According to an agreement between the Mathematical Association of America and the editors of the Annals of Mathematics, the Association contributes to the support of the Annals, and the Annals is supplied to individual members of the Association at one half of the regular price. In consequence of this agreement the volume of the Annals was increased by 100 pages, which are devoted to expository and historical articles in so far as suitable articles of this class are obtainable. Thus far the editors have not received enough such articles to fill the space available, and therefore wish to call the attention of authors to this lack and to the fact that, as long as the shortage continues, expository or historical articles of sufficient merit will receive prompt publication.

A number of the expository articles which have already been published are available in separate form and are listed for sale on the inside of the back cover of this number of the Annals. The regular subscription price of the Annals is $3.00 a volume.

Entered at the postoffice at Lancaster, Pa., as second-class matter under the Act of March 3, 1879.

Copies of the following memoirs can be obtained by addressing THE ANNALS OF MATHEMATICS, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. JENSEN. Authorized translation with additional notes by T. H. GRONWALL. 43 pages. Price 50 cents.

The gamma function in the integral calculus. By T. H. GRONWALL. 89 pages. 90 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. DICKSON. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. OSGOOD. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. JENSEN. Authorized translation from the Danish by T. H. GRONWALL. 29 pages. Price 40 cents.

CONTENTS

Page

Dirichlet's problem. By G. E. RAYNOR 183
Annihilators of modular invariants and convariants. By OLIVE C. HAZLETT 198
Systems of linear inequalities. By W. B. CARVER 212
Euler squares. By H. F. MACNEISH 221
Geometric aspects of Einstein's theory. By J. PIERPONT 228
Cauchy's paper of 1814 on definite integrals. By H. J. ETTLINGER . . . 255
Arithmetical deduction of Kronecker's class-number relations. By G. H. CRESSE 271
Cyclotomic heptasection for the prime 43. By PANDIT OUDH UPADHYAYA 280
Summation of a double series. By T. H. GRONWALL 282

ANNALS OF MATHEMATICS

Published in September, December, March and June at Lancaster, Pa., and Princeton, N. J. Subscription price, $3 a volume (four numbers) in advance. Single copies $1.00. Subscriptions, orders for back numbers, and changes of address should be sent to the PRINCETON UNIVERSITY PRESS, PRINCETON, NEW JERSEY.

Manuscripts and all editorial correspondence should be addressed to THE ANNALS OF MATHEMATICS, P. O. BOX 53, PRINCETON, NEW JERSEY. Manuscripts should be typewritten, with the exception of formulæ, and must be in final form with all references filled in.

Authors receive gratis 25 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

PRESS OF
THE NEW ERA PRINTING COMPANY
LANCASTER, PA.

Stanford Library
JUL 25 1923
Math

JUNE, 1922

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

ORMOND STONE — J. W. ALEXANDER
L. P. EISENHART — T. H. GRONWALL
OSWALD VEBLEN — J. H. M. WEDDERBURN

WITH THE COÖPERATION OF

A. A. BENNETT — H. BLUMBERG
G. A. PFEIFFER — J. K. WHITTEMORE

PUBLISHED BY THE

PRINCETON UNIVERSITY PRESS

SECOND SERIES, VOL. 23, NO. 4

LANCASTER, PA., AND PRINCETON, N. J.

1923

According to an agreement between the Mathematical Association of America and the editors of the Annals of Mathematics, the Association contributes to the support of the Annals, and the Annals is supplied to individual members of the Association at one half of the regular price. In consequence of this agreement the volume of the Annals was increased by 100 pages, which are devoted to expository and historical articles in so far as suitable articles of this class are obtainable. Thus far the editors have not received enough such articles to fill the space available, and therefore wish to call the attention of authors to this lack and to the fact that, as long as the shortage continues, expository or historical articles of sufficient merit will receive prompt publication.

A number of the expository articles which have already been published are available in separate form and are listed for sale on the inside of the back cover of this number of the Annals. The regular subscription price of the Annals is $3.00 a volume.

Entered at the postoffice at Lancaster, Pa., as second-class matter under the Act of March 3, 1879.

Copies of the following memoirs can be obtained by addressing THE ANNALS OF MATHEMATICS, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. JENSEN. Authorized translation with additional notes by T. H. GRONWALL. 43 pages. Price 50 cents.

The gamma function in the integral calculus. By T. H. GRONWALL. 89 pages. 90 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. DICKSON. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. OSGOOD. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. JENSEN. Authorized translation from the Danish by T. H. GRONWALL. 29 pages. Price 40 cents.

According to an agreement between the Mathematical Association of America and the editors of the Annals of Mathematics, the Association contributes to the support of the Annals, and the Annals is supplied to individual members of the Association at one half of the regular price. In consequence of this agreement the volume of the Annals was increased by 100 pages, which are devoted to expository and historical articles in so far as suitable articles of this class are obtainable. Thus far the editors have not received enough such articles to fill the space available, and therefore wish to call the attention of authors to this lack and to the fact that, as long as the shortage continues, expository or historical articles of sufficient merit will receive prompt publication.

A number of the expository articles which have already been published are available in separate form and are listed for sale on the inside of the back cover of this number of the Annals. The regular subscription price of the Annals is $3.00 a volume.

Entered at the postoffice at Lancaster, Pa., as second-class matter under the Act of March 3, 1879.

Copies of the following memoirs can be obtained by addressing The Annals of Mathematics, Princeton, N. J.:

An elementary exposition of the theory of the gamma function. By J. L. W. V. Jensen. Authorized translation with additional notes by T. H. Gronwall. 43 pages. Price 50 cents.

The gamma function in the integral calculus. By T. H. Gronwall. 89 pages. 90 cents.

Fermat's last theorem and the origin and nature of the theory of algebraic numbers. By L. E. Dickson. 27 pages. Price 35 cents.

Factorization of analytic functions of several variables. By W. F. Osgood. 19 pages. Price 25 cents.

Investigation of a class of fundamental inequalities in the theory of analytic functions. By J. L. W. V. Jensen. Authorized translation from the Danish by T. H. Gronwall. 29 pages. Price 40 cents.

CONTENTS

Page

On the positions of the imaginary points of inflexion and critic centers of a real cubic. By B. M. TURNER 287
Frequency distributions obtained by certain transformations of normally distributed variates. By H. L. RIETZ 292
The associated point of seven points in space. By H. S. WHITE 301
Common solutions of two simultaneous Pell equations. By A. ARWIN . . . 307
On the complete independence of Hurwitz's postulates for abelian groups and fields. By B. A. BERNSTEIN 313
On power series with positive real part in the unit circle. By T. H. GRONWALL 317
Algebraic surfaces, their cycles and integrals. A correction. By S. LEFSCHETZ 333

ANNALS OF MATHEMATICS

Published in September, December, March and June at Lancaster, Pa., and Princeton, N. J. Subscription price, $3 a volume (four numbers) in advance. Single copies $1.00. Subscriptions, orders for back numbers, and changes of address should be sent to the PRINCETON UNIVERSITY PRESS, PRINCETON, NEW JERSEY.

Manuscripts and all editorial correspondence should be addressed to THE ANNALS OF MATHEMATICS, P. O. BOX 53, PRINCETON, NEW JERSEY. Manuscripts should be typewritten, with the exception of formulæ, and must be in final form with all references filled in.

Authors receive gratis 25 reprints of each article, postage prepaid. Additional copies will be furnished at cost.

LANCASTER PRESS, INC.
LANCASTER, PA.